



CENTRO DE INVESTIGACIONES BIOLÓGICAS  
DEL NOROESTE, S.C.

---

Programa de Estudios de Posgrado

ESTRATEGIAS DE TRAZABILIDAD GENÉTICA Y ANÁLISIS DE  
PARENTESCO EN CAMARÓN BLANCO *Penaeus (Litopenaeus)*  
*vannamei*.

T E S I S

Que para obtener el grado de

Doctora en Ciencias

Uso, Manejo y Preservación de los Recursos Naturales  
(Orientación en Acuicultura)

P r e s e n t a

**Adriana Max Aguilar**

La Paz, Baja California Sur, mayo de 2024.

## ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 11:30 horas del día 06 del Mes de mayo del 2024, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado y Formación de Recursos Humanos del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

**"Estrategias de trazabilidad genética y análisis de parentesco en camarón blanco *Penaeus (Litopenaeus) vannamei*"**

Presentada por el alumno:

**Adriana Max Aguilar**

Aspirante al Grado de DOCTOR EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN Acuicultura

Después de intercambiar opiniones los miembros de la Comisión manifestaron su **APROBACIÓN DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

### LA COMISIÓN REVISORA



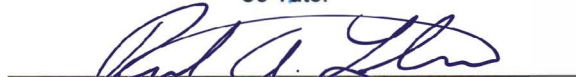
Dr. Ricardo Pérez Enríquez  
Director de Tesis



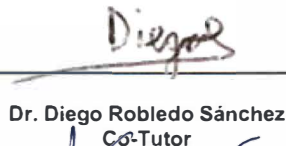
Dra. Ana María Ibarra Humphries  
Co-Tutor



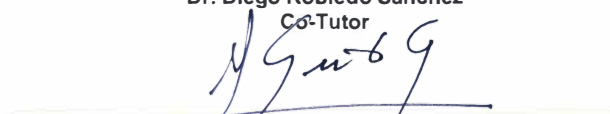
Dr. Jorge Hernández López  
Co-Tutor



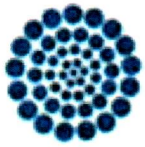
Dr. Raúl Antonio Llera Herrera  
Co-Tutor



Dr. Diego Robledo Sánchez  
Co-Tutor



Dra. Alejandra Nieto Garibay,  
Directora de Estudios de Posgrado y  
Formación de Recursos Humanos



**CONAHCYT**  
CONSEJO NACIONAL DE HUMANIDADES  
CIENCIAS Y TECNOLOGÍAS

**CI3** Centro de  
Investigaciones  
Biológicas  
del Noroeste

La Paz, Baja California Sur, a 03 de mayo de 2024.

Los miembros del comité de tesis del (la) estudiante Adriana Max Aguilar del Programa de Doctorado en Ciencias en el Uso, Manejo y Preservación de los Recursos Naturales revisamos el contenido de la tesis y otorgamos el Vo.Bo. dado que la tesis no representa un plagio de otro documento como lo muestra el reporte de similitud realizado:

- Herramienta antiplagio:  
iThenticate
- Filtros utilizados:  
Excluir citas  
Excluir bibliografía
- Porcentajes de similitud:  
13% Se muestra captura de pantalla

openaeus) Citas excluidas 13%  
Bibliografía excluida

RESUMEN DE COINCIDENCIAS

Rank	Source	Words	Percentage
1	Internet Copiado el 15-Oct-2022 cibnor.repositorioinstitucional.mx	296 palabras	2%
2	Internet hdl.handle.net	262 palabras	1%
3	Internet Copiado el 15-Oct-2022 cicese.repositorioinstitucional.mx	186 palabras	1%
4	Internet Copiado el 27-Nov-2020 doczz.es	84 palabras	<1%
5	Internet Copiado el 22-Dic-2019 www.jove.com	63 palabras	<1%
6	Internet Copiado el 25-Mar-2021 www.scielo.org.mx	53 palabras	<1%
7	Internet Copiado el 25-Sep-2022 repository.javeriana.edu.co	49 palabras	<1%
-	Internet	46 palabras	

**Firmas del Comité**

Dr. Ricardo Pérez Enríquez

Dra. Ana María Ibarra Humphries

Dr. Jorge Hernández López

Dr. Raúl A. Llera Herrera

Dr. Diego Robledo Sánchez

## **Conformación de Comités**

### **Comité Tutorial**

Dr. Ricardo Pérez Enríquez  
Centro de Investigaciones Biológicas del Noroeste, S. C.  
Director de Tesis

Dra. Ana María Ibarra Humphries  
Centro de Investigaciones Biológicas del Noroeste, S. C.  
Co-Tutora de Tesis

Dr. Jorge Hernández López  
Centro de Investigaciones Biológicas del Noroeste, S. C.  
Co-Tutor de Tesis

Dr. Raúl Antonio Llera Herrera  
Instituto de Ciencias del Mar y Limnología. UNAM  
Co-Tutor de Tesis

Dr. Diego Robledo Sánchez  
Instituto Roslin. Universidad de Edimburgo  
Co-Tutor de Tesis

### **Comité Revisor de Tesis**

Dr. Ricardo Pérez Enríquez  
Dra. Ana María Ibarra Humphries  
Dr. Jorge Hernández López  
Dr. Raúl Antonio Llera Herrera  
Dr. Diego Robledo Sánchez

### **Jurado de Examen**

Dr. Ricardo Pérez Enríquez  
Dra. Ana María Ibarra Humphries  
Dr. Jorge Hernández López  
Dr. Raúl Antonio Llera Herrera  
Dr. Diego Robledo Sánchez

### **Suplentes**

Dra. Cristina Escobedo Fregoso  
Dr. Pedro Cruz Hernández

## Resumen

El camarón blanco *Penaeus vannamei* es la especie acuícola más importante en México por su valor comercial. La tasa crecimiento en los últimos 10 años fue del 4.56% sin embargo, su producción ha sufrido afectaciones importantes, principalmente por la presencia de enfermedades. La implementación de programas de mejoramiento genético puede incrementar la producción a través del desarrollo de líneas seleccionadas genéticamente. Estos programas requieren llevar el seguimiento del pedigrí de los individuos para definir esquemas de entrecruzamiento selectivos y evitar la pérdida de variabilidad, ya que deficiencias en el manejo pueden llevar a la depresión por endogamia, afectando el desempeño del cultivo. Por otro lado, la caracterización genética de líneas de cultivo permite conocer su estado genético, así como la posible trazabilidad y autenticación de origen de los productos acuícolas, esto es importante por cuestiones de seguridad alimentaria, así como para evitar prácticas fraudulentas. El objetivo de este trabajo fue la caracterización genómica de poblaciones de cultivo y silvestres en México, así como la evaluación de estrategias de trazabilidad y análisis de parentesco en *P. vannamei* mediante el desarrollo e implementación de paneles de SNPs (polimorfismos de una base) de baja densidad. Se desarrollaron paneles de SNPs con las técnicas 2bRAD y GT-Seq. La caracterización genómica de lotes de cultivo con este último permitió distinguir hasta seis grupos genéticos con diferentes orígenes de importación en México, y una clara diferenciación genética entre un lote de cultivo con origen en México y los lotes con origen de importación; además, se observó similitud entre perfiles genéticos de algunos lotes de cultivo y la muestra silvestre, lo que indicaría la posible introgresión de organismos de cultivo a las poblaciones naturales. En la validación del panel de SNPs para trazabilidad, se logró la asignación del 46% de muestras “ciegas” a cinco grupos genéticos de origen conocido; la falta de asignación del resto de los individuos se debió a la ausencia de muestras de referencia para algunos orígenes (como el lote de importación de Hawái), así como a cambios en las frecuencias alélicas por los diferentes tipos de manejo dentro y entre los laboratorios de producción. Sin embargo, este panel puede ser utilizado para verificar el origen declarado de un lote de cultivo con la presunta población de origen. El panel de SNPs también fue validado para análisis de parentesco con hasta el 100% de asignación de la progenie a ambos padres, así como con la asignación correcta de las familias esperadas; esto, aunado a la estandarización de genotipificación enfocada a reducir el tiempo y costo de análisis, lo convierte en una buena opción para los productores de camarón que requieren de este tipo de análisis. Este panel puede cumplir tres objetivos en una sola reacción multiplex con la misma plataforma de obtención de genotipos (GT-seq): 1) para estudiar la estructura genética y la asignación de muestras a una población de referencia, 2) para análisis de parentesco, y 3) identificación de sexo. Por otra parte, el panel de SNPs desarrollado con 2bRAD es una herramienta para futuros análisis que requieran paneles de alta densidad, como estudios de asociación a lo largo del genoma (GWAS) y Selección Genómica en *P. vannamei*, los cuales permitirían conocer la arquitectura genética de caracteres de interés, así como optimizar estrategias selección.

**Palabras clave:** SNP, trazabilidad, análisis de parentesco, GT-seq, *P. vannamei*

**ORCID:** 0000-0001-5955-5957



Vo.Bo. Dr. Ricardo Pérez Enríquez  
Director de Tesis

## Summary

The whiteleg shrimp *Penaeus vannamei* is the main aquaculture species in Mexico due to its commercial value. The growth rate in the last ten years was 4.56%. However, its production has been affected mainly by diseases. The implementation of breeding programs is an effective approach to enhance aquaculture production by developing improved genetic stocks. These programs require monitoring of the pedigree to define selective breeding schemes and avoid the loss of variability through generations, because management deficiencies lead to inbreeding depression, affecting the organisms' performance. On the other hand, the genetic characterization of aquaculture stocks allows knowing their genetic status, as well as the possible traceability and authentication of the origin of aquaculture products, this is important for food safety, as well as to avoid fraudulent practices. The objective of this work was the genomic characterization of aquaculture and wild stocks in Mexico, as well as evaluate traceability strategies and parentage analysis in *P. vannamei* through low-density SNP (single base polymorphisms) panels. SNP panels were developed by 2bRAD and GT-Seq techniques. The genomic characterization of cultured hatcheries with the last one allowed the distinction of up to six genetic groups with different import origins in Mexico, and a clear genetic differentiation between a hatchery from Mexico and commercially important lines. Furthermore, similarity was observed between the genetic profiles of some cultured stocks and wild populations, which could indicate the possible introgression of aquaculture organisms into the environment. In the validation of the low-density SNP panel for traceability, the assignment of 46% of the "blind" samples to the five genetic groups of known origin previously reported for Mexico was achieved. The lack of assignment of the rest of the individuals was due to the lack of reference samples for some origins (as in the case of the Hawaii-imported batch), and changes in allele frequencies due to different types of management within and between hatcheries. Nevertheless, this panel can be used to verify the declared origin of a batch with the presumed population of origin. The SNP panel was also validated for parentage analysis with up to 100% assignment of progeny to both parents, as well as with correct assignment of expected families; this, combined with the standardization of genotyping focused on reducing time and cost of analysis, makes it a good option for shrimp producers who require these analyzes. Therefore, this panel can fulfill three objectives in a single multiplex reaction with the same genotyping platform (GT-seq): 1) for genetic structure and assignment of samples to a reference population, 2) for parentage analysis, and 3) sex identification. On the other hand, the SNP panel developed with 2bRAD is a tool for future analyses that require a high-density SNP panel, such as genome-wide association studies (GWAS) and Genomic Selection in *P. vannamei*, which would allow to know the genetic architecture of economic traits of interest, as well as optimizing selection strategies.

**Keywords:** SNP, traceability, parentage analysis, GT-seq, *P. vannamei*

**ORCID:** 0000-0001-5955-5957



Vo.Bo. Dr. Ricardo Pérez Enríquez  
Director de Tesis

**Dedicatoria**

*A mi Papá, Julio R. Max, quién desafortunadamente no pudo llegar a la meta conmigo, pero sé que hubiera estado en primera fila para la culminación de mi doctorado (y apurándome también para concluir esta tesis), con sus grandes virtudes y defectos, ¡Te amo y te extraño!*

*A mi Abuela Ramona Malpica, quién era una gran mujer, que con su gran fortaleza y valentía superó tantas dificultades en su vida y salió adelante en una época en donde las oportunidades para las mujeres eran escasas o nulas. ¡Gracias por haber sido parte de mi vida y un ejemplo a seguir!*

*A mi Esposo Luis Manuel Hernández, por estar conmigo en las buenas y en las malas, por acompañarme durante este trayecto que resultó mucho más difícil de lo que esperábamos.*

*A mi Mamá, Ruth Aguilar, mi Hermana, Wendy Max y mi Cuñado José Augusto.*

*Por último, y no menos importante, a mi Sobrina Kikey, quién con su llegada a la familia nos llenó de felicidad a todos.*

## **Agradecimientos**

Al Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCYT) por la beca de Doctorado otorgada 413201 y por el financiamiento del proyecto Fondo CONACYT-Ciencia de Frontera 2021 (319686).

Al Consejo Sudcaliforniano de Ciencia y Tecnología (COSCYT) por el financiamiento del proyecto “Plataforma tecnológica de análisis genético multiespecie de última generación para potenciar el desarrollo acuícola de Baja California Sur”, número de convenio CAR/14/2021.

Al Centro de Investigaciones Biológicas del Noroeste S. C. por aceptarme en el programa de Posgrado como alumna de Doctorado y por todas las facilidades prestadas.

Agradezco a todo personal de Posgrado, por todo su apoyo, en especial, durante la pandemia, que fue una época difícil para todos los alumnos, a la Mtra. Osvelia Ibarra Morales, Tania Verónica Núñez Valdez, Horacio Sandoval Gómez, así como al personal de Becas y apoyos. Al personal de la biblioteca por su trato siempre cordial, Susana Luna García, Ana María Talamantes Cota y María Esther Ojeda Castro.

A la empresa Genamex y al Dr. Francisco Magallón por las muestras de camarón proporcionadas, así como a Bluegenetics México y a su personal, en especial a Melisa y Christian. Al comité de Sanidad Acuícola de Guerrero y a Fidencio Hernández por las muestras silvestres de Guerrero. Al Laboratorio de Genómica y Bioinformática del CIBNOR, un enorme agradecimiento a Gabriela Mendoza Carrión, a la Dra. Cristina Escobedo y al Dr. Ricardo Vázquez Juárez.

Al Laboratorio de Fisiología Comparada, en especial a Paty Hinojosa, quien con mucha amabilidad y una sonrisa siempre apoyaba a los estudiantes que utilizábamos el Laboratorio (Q.E.P.D).

Al Laboratorio Húmedo de Nutrición Acuícola por permitirme el uso de sus instalaciones y el apoyo otorgado, a Sandra de la Paz Reyes y Gabriel Robles Villegas.



Al Área de Servicios e Infraestructura Informática (Redes), a María Isabel Castro Hernández y Jorge Mario Rodríguez Meza, por las facilidades para el uso del servidor.

Durante el transcurso del Doctorado, tuve la fortuna de que me tocaran dos excelentes directoras de Posgrado, a quienes agradezco enormemente. A la Dra. Gracia Alicia Gómez Anduro, por todo su apoyo y comprensión durante la pandemia, quién se preocupó por apoyar a los alumnos no sólo en los aspectos académicos, sino también en aspectos de salud mental y personal, aspectos tan importantes y que muchas veces no tomamos en cuenta para concluir con éxito nuestros estudios. Agradezco a la Dra. Alejandra Nieto Garibay por su comprensión y apoyo invaluable, por tomarse el tiempo de platicar conmigo que, aunque fueran unos minutos, fueron parte clave para que continuara y concluyera este último trayecto.

Al área de acompañamiento Psicológico, a la Psic. Refugio Castro Argüelles, por escucharme y ser una red de apoyo en ese momento tan difícil de mi vida, por las herramientas brindadas.

Un enorme agradecimiento a mi Director de Tesis, el Dr. Ricardo Pérez Enríquez. Gracias por estos casi 13 años de colaboración, que han sido invaluable, que me han permitido crecer como profesional. Pero, sobre todo, no me alcanzan las palabras para agradecerle por estos últimos años que han sido más que difíciles, gracias por su confianza, su comprensión, por seguir apoyándome y creyendo en mí cuando ya ni siquiera yo creía en mí, pero usted de alguna manera supo que iba a lograr superar todos esos obstáculos. De verdad, que no me alcanzan las palabras. Agradezco a mi Comité Tutorial: la Dra. Ana María Ibarra Humphries, el Dr. Diego Robledo, el Dr. Raúl Llera y al Dr. Jorge Hernández López. Gracias por todas sus contribuciones y su apoyo durante el desarrollo de este trabajo, no pude tener mejor comité tutorial, conformado por excelentes investigadores a quienes escuchar y admirar. Muchas gracias por su comprensión que, aunque todo este proyecto resultó muy diferente, continuaron apoyándome, cuando otros investigadores ya no lo hubieran hecho.

Al Laboratorio de Genética Acuícola; un especial agradecimiento a la M.C. Susana Ávila, no sólo por el apoyo brindado durante estos años en el Laboratorio, si no por su valiosa amistad y por esas pláticas, muchas veces tan atinadas y que hacían falta para tomar un respiro y continuar.

Agradezco a mis compañeros del Laboratorio, en especial a la Dra. Rosa María Morelos Castro, a Christopher y Pavel por el apoyo brindado.

Aunque esta tesis está dedicada a ellos, reitero mi agradecimiento a mi familia por su apoyo, en especial a mi Esposo, el Biol. Luis Manuel Hernández, no sólo por el apoyo personal y emocional, si no por su apoyo en el mantenimiento y muestreo de los camarones; por ser un gran compañero de vida.

## Contenido

<b>Resumen</b> .....	i
<b>Summary</b> .....	ii
<b>Dedicatoria</b> .....	iii
<b>Agradecimientos</b> .....	iv
<b>Contenido</b> .....	vii
<b>Lista de figuras</b> .....	ix
<b>Lista de tablas</b> .....	xi
<b>1.INTRODUCCIÓN</b> .....	1
<b>2.ANTECEDENTES</b> .....	4
2.1 Marcadores moleculares y técnicas de obtención de genotipos .....	4
2.2 Caracterización genética de poblaciones de cultivo y silvestres .....	6
2.3 Trazabilidad y asignación poblacional .....	8
2.4 Análisis de pedigrí en la acuicultura .....	9
<b>3.JUSTIFICACIÓN</b> .....	11
<b>4.HIPÓTESIS</b> .....	13
<b>5.OBJETIVOS</b> .....	14
5.1 Objetivo general .....	14
5.2 Objetivos particulares .....	14
<b>6.MATERIAL Y MÉTODOS</b> .....	15
6.1 Desarrollo de paneles de marcadores moleculares tipo SNP en camarón blanco <i>P. vannamei</i> .....	15
6.1.1 Obtención de panel de SNPs <i>de novo</i> con 2bRAD .....	15
6.1.1.1 Evaluación <i>in silico</i> de enzimas de restricción tipo IIB .....	15
6.1.1.2 Estandarización de 2bRAD para obtención de panel de SNPs.....	18
6.1.2 Obtención de panel SNPs utilizando datos públicos de secuenciación.....	21
6.1.2.1 Análisis bioinformático para obtención de SNPs .....	21
6.1.2.2 Análisis <i>in silico</i> para evaluar reacción multiplex para la obtención de genotipos por GT-seq.....	22
6.2 Caracterización genómica de poblaciones de cultivo de <i>Penaeus vannamei</i> en México con un panel de baja densidad de SNPs y validación para análisis de asignación poblacional .....	23
6.2.1 Evaluación <i>in silico</i> de panel de SNPs para asignación poblacional y estructura genética .....	23
6.2.2 Estandarización de obtención de genotipos mediante curvas de disociación de alta resolución (HRM). .....	24
6.2.3 Caracterización genómica de poblaciones de cultivo y análisis de asignación poblacional.....	26
6.3 Validación de panel de SNPs de baja densidad para análisis de parentesco en cultivo de camarón blanco <i>P. vannamei</i> .....	27
6.3.1 Evaluación <i>in silico</i> de panel de SNPs para análisis de parentesco .....	27
6.3.2 Estandarización de GT-seq para obtención de genotipos.....	28
6.3.3 Validación de panel de SNPs para análisis de parentesco .....	31
<b>7.RESULTADOS</b> .....	33
7.1 Desarrollo de paneles de marcadores moleculares tipo SNP en camarón blanco <i>P. vannamei</i> .....	33

7.1.1 Panel de SNPs con 2bRAD .....	33
7.1.2 Panel de SNPs utilizando datos públicos de secuenciación .....	35
7.1.2.1 Análisis bioinformático para obtención de SNPs .....	35
7.1.2.2 Análisis <i>in silico</i> para evaluar reacción multiplex .....	36
7.2 Caracterización genómica de poblaciones de cultivo de <i>P. vannamei</i> en México con un panel de baja densidad de SNPs y validación para análisis de asignación poblacional .....	37
7.2.1 Análisis <i>in silico</i> para validar panel de SNPs para asignación poblacional y estructura genética.....	37
7.2.2 Estandarización de obtención de genotipos mediante HRM.....	39
7.2.3 Caracterización genómica de poblaciones de cultivo y análisis de asignación poblacional.....	41
7.2.4 Evaluación <i>in silico</i> de panel de SNPs para análisis de parentesco .....	45
7.2.5 Obtención de genotipos por GT-seq.....	46
7.2.6 Validación de panel de SNPs para análisis de parentesco .....	48
<b>8.DISCUSIÓN</b> .....	49
8.1 Panel de SNPs con 2bRAD .....	49
8.2 Panel de 118 SNPs y obtención de genotipos con GT-seq.....	50
8.3 Caracterización genómica de poblaciones de cultivo de <i>P. vannamei</i> en México y validación de panel de SNPs de baja densidad para análisis de asignación poblacional y trazabilidad....	51
8.4 Validación de panel de SNPs de baja densidad para análisis de parentesco en cultivo de camarón blanco <i>Penaeus vannamei</i> .....	53
<b>9.CONCLUSIONES</b> .....	56
<b>10.LITERATURA CITADA</b> .....	58
<b>11.ANEXOS</b> .....	69
Anexo A. Comandos utilizados para identificar los sitios de reconocimiento de enzimas IIB .	69
Anexo B. Secuencias de adaptadores y oligos para preparación de librerías 2bRAD .....	72
Anexo C. Descripción de análisis de datos generados durante la qPCR-HRM para curvas de alta resolución (HRM) y obtención de genotipos de SNPs .....	74
Anexo D. Temperatura de alineamiento y secuencia de oligos para HRM del panel de SNPs para análisis de estructura genética y asignación poblacional.....	75
Anexo E. Panel de 118 SNPs para estructura genética y asignación poblacional, análisis de parentesco e identificación de sexo.....	78
Anexo F. Secuencia de oligos para preparación de librerías de GT-seq .....	86
Anexo G. Comandos de Perl para obtención de genotipos de librerías GT-seq .....	88
Anexo H. Archivo de entrada para obtención de genotipos en Perl de librerías GT-seq.....	93
Anexo I. Costo de genotipificación de panel de 118 SNPs con GT-seq.....	99
Anexo J. Artículo Aceptado.-Caracterización genética de especies acuícolas mediante paneles de SNPs de baja densidad. ....	101
Anexo K. Artículo Sometido.-Advances in genetic and genomic resources in Pacific whiteleg shrimp <i>Penaeus (Litopenaeus) vannamei</i> - towards modern and sustainable shrimp aquaculture .....	120

## Lista de figuras

<b>Figura 1.</b> Volumen de producción acuícola de <i>P. vannamei</i> en México 2007-2021. Datos reportados por CONAPESCA (CONAPESCA, 2017; CONAPESCA, 2021). .....	1
<b>Figura 2.</b> Sitios de reconocimiento y corte de enzimas tipo IIB. En azul, el sitio de reconocimiento de cada enzima. En rojo, las bases nucleotídicas que corresponden a los extremos cohesivos y que, en los adaptadores, pueden ser substituidas por bases selectivas. ....	17
<b>Figura 3.</b> Fragmentos de ADN después de la ligación de adaptadores. Adaptadores de doble cadena en color verde. A) Ligación de adaptador 1 con bases selectivas TC y adaptador 2 sin bases selectivas. B) Ligación de adaptadores 1 y 2 con bases selectivas. ....	18
<b>Figura 4.</b> Visualización de librerías 2bRAD en gel de agarosa al 2% previo a la purificación de fragmentos. El fragmento esperado de 180 pb se señala en el rectángulo rojo. ....	20
<b>Figura 5.</b> Visualización de fragmentos de PCR1 en gel de agarosa al 2%. Fragmentos esperados de 126-216 pb señalados en el rectángulo rojo. ....	29
<b>Figura 6.</b> Visualización de fragmentos de PCR2 previo a la normalización y purificación. Fragmentos esperados de 191-281 pb resaltados en el rectángulo rojo. ....	30
<b>Figura 7.</b> Análisis Discriminante de Componentes Principales (DAPC). A) DAPC con 35 SNPs. B) DAPC con 21 SNPs. Pop1Tx (origen Texas), Pop2Ec (origen Ecuador), Pop3CA1 (origen Centroamérica 1), Pop4Mx (origen México), Pop5CA2 (origen Centroamérica 1). Grupos genéticos reportados por Perez-Enriquez <i>et al.</i> (2018c). ....	38
<b>Figura 8.</b> Análisis Discriminante de Componentes Principales (DAPC) con 117 SNPs. Pop1Tx (origen Texas), Pop2Ec (origen Ecuador), Pop3CA1 (origen Centroamérica 1), Pop4Mx (origen México), Pop5CA2 (origen Centroamérica 1). Grupos genéticos reportados por Perez-Enriquez <i>et al.</i> (2018c). ....	39
<b>Figura 9.</b> Curvas HRM de SNP Lv24611_43_128. A) Curvas de disociación; se señala con las líneas segmentadas los puntos en los que se delimita la curva para su posterior análisis. B) Curva derivativa en el que se observan homocigotos para ambos alelos. ....	40
<b>Figura 10.</b> Ejemplo de curvas derivativas. Las líneas del mismo color corresponden a un mismo individuo analizado por triplicado. A) SNP donde no se distingue claramente la diferencia entre genotipos. B) SNP donde se observan 4 genotipos distintos. ....	40
<b>Figura 11.</b> Gráfica de DeltaK para visualizar el número de K más probable (K=grupos genéticos) incluyendo los 14 lotes de cultivo (muestras ciegas) y las muestras de referencia (Perez-Enriquez <i>et al.</i> , 2018c). ....	41
<b>Figura 12.</b> Gráfica de banderas de Structure para K=3 y K=6. Muestras “ciegas” de 14 lotes de cultivo (L), entre paréntesis origen reportado por el laboratorio de producción; muestras de origen conocido (Pop) reportados por Perez-Enriquez <i>et al.</i> (2018c). Bvis=Buenavista, CAI=Centroamérica1, CAII=Centroamérica2, Ec=Ecuador, Hw=Hawái, Mx=México, Nc=Nicaragua, Tx=Texas. ....	42
<b>Figura 13.</b> Gráfica de DeltaK para visualizar el número de K más probable (K=grupos genéticos) incluyendo los 14 lotes de cultivo (muestras ciegas), 3 poblaciones silvestres (Sinaloa, Sinaloa Norte y Guerrero) y las muestras de referencia (Perez-Enriquez <i>et al.</i> , 2018c). ....	44
<b>Figura 14.</b> Gráfica de banderas de Structure para K=2 y K=7. Muestras “ciegas” de 14 lotes de cultivo (L), entre paréntesis origen reportado por el laboratorio de producción; muestras silvestres (W); muestras de origen conocido (Pop) reportados por Pérez-Enriquez <i>et al.</i> (2018c). Bvis=Buenavista, CAI=Centroamérica1, CAII=Centroamérica2, Ec=Ecuador, Hw=Hawái,	

Mx=México, Nc=Nicaragua, Tx=Texas, WSin=Sinaloa, WGue=Guerrero, WNs=Sinaloa noroeste.  
..... 44

**Figura 15.** Gráfica de DAPC (Análisis Discriminante de Componentes Principales) para las muestras “ciegas” de 14 lotes de cultivo (L); muestras silvestres (W), Sin=Sinaloa, Gue=Guerrero, Ns=Sinaloa noroeste; muestras de origen conocido (Pop) reportados por Pérez-Enriquez *et al.* (2018c), Pop1Tx=Texas, Pop2Ec=Ecuador, Pop3CA1=Centroamérica1, Pop4Mx=México, Pop5CA2=Centroamérica2. A) DAPC de Componentes Principales 1 y 2. B) DAPC de Componentes Principales 1 y 3..... 45

**Figura 16.** Promedio de lecturas objetivo para cada locus del panel de 118 SNPs..... 47

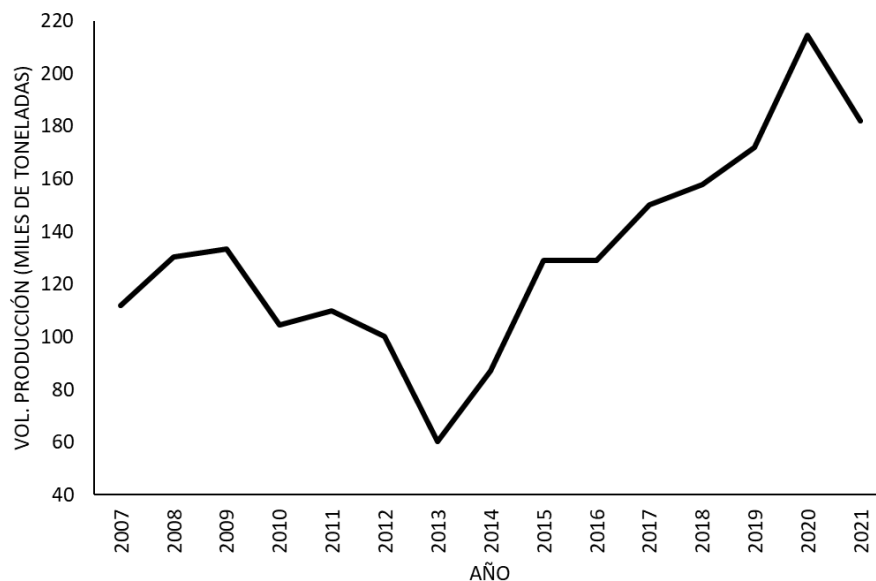
**Lista de tablas**

<b>Tabla 1.</b> Secuencias de reconocimiento y corte de enzimas de restricción tipo IIB y tamaño del fragmento obtenido. ....	16
<b>Tabla 2.</b> Número de sitios de reconocimiento de enzimas IIB en genoma de camarón .....	33
<b>Tabla 3.</b> Porcentaje de fragmentos remanentes al incluir adaptadores con bases selectivas .....	34
<b>Tabla 4.</b> Número de lecturas y porcentaje de alineamiento de librerías 2bRAD .....	35
<b>Tabla 5.</b> Asignación poblacional con diferente número de SNPs en análisis <i>in silico</i> .....	37
<b>Tabla 6.</b> Origen reportado de lotes de cultivo por laboratorios de producción y origen asignado con análisis de Structure (asignación de grupos genéticos) y GeneClass (asignación de individuos). .....	43
<b>Tabla 7.</b> Porcentaje de asignación en análisis de parentesco <i>in silico</i> con 117 SNPs .....	46
<b>Tabla 8.</b> Porcentajes de asignación de progenie a ambos padres (n=80) .....	48

## 1. INTRODUCCIÓN

La acuicultura es una actividad económicamente importante para satisfacer la demanda de proteína animal a nivel mundial. De la producción de especies animales acuáticas del 2021, la acuicultura representó el 50% con 90.9 millones de toneladas con un valor comercial de 281 mil millones USD (FAO, 2024). De esta producción, el camarón blanco *Penaeus (Litopenaeus) vannamei* es una de las especies más importantes; representa el 53% de la producción de crustáceos y el 6.9% de la producción acuícola total con 6.3 millones de toneladas y un valor comercial aproximado de 36.5 mil millones USD (FAO, 2024).

En México, *P. vannamei* ocupa el primer lugar por su valor comercial con una producción de 182,000 toneladas en 2021, y un valor aproximado de \$17,314 millones de pesos (CONAPESCA, 2021). La tasa crecimiento en los últimos 10 años fue del 4.56% (CONAPESCA, 2021) sin embargo, la producción ha sufrido importantes afectaciones, como en 2013 (Fig.1) debido a la presencia de enfermedades, como el Síndrome de la Mancha Blanca (WSD) causada por el Virus del Síndrome de la Mancha blanca (WSSV) y la Enfermedad de la Necrosis Hepatopancreática Aguda (AHPND), cuyo agente causante se ha reportado en México como *Vibrio parahaemolyticus* (Soto-Rodriguez *et al.*, 2015), específicamente aquellas cepas que contienen un plásmido, con genes que codifican toxinas tipo pirA y pirB (Han *et al.*, 2015; Lee *et al.*, 2015).



**Figura 1.** Volumen de producción acuícola de *P. vannamei* en México 2007-2021. Datos reportados por CONAPESCA (CONAPESCA, 2017; CONAPESCA, 2021).



Los programas de mejoramiento genético pueden contribuir al incremento de la producción a través del desarrollo de líneas genéticamente mejoradas (Gjedrem *et al.*, 2012), en las cuales las características seleccionadas son acumulativas y permanentes (Gjedrem y Baransky, 2009). Para ello, la diversidad genética es necesaria para optimizar los caracteres económicamente importantes, como crecimiento, supervivencia, resistencia a enfermedades, entre otras (Davis y Hetzel, 2000; Gjedrem y Baranski, 2009). El seguimiento del pedigrí de los individuos es esencial para definir esquemas selectivos de entrecruzamiento y evitar la pérdida de variabilidad a través de las generaciones de cultivo, deficiencias en el manejo pueden llevar a un incremento en la endogamia y a la depresión por endogamia (Gjerde *et al.*, 1996; Gjerde y Rye, 1998; Doyle *et al.*, 2001).

Por otro lado, la caracterización e identificación genética de líneas de cultivo permiten conocer su estado genético, así como la trazabilidad y autenticación de origen de los productos acuícolas, lo cual ha cobrado importancia por cuestiones de seguridad alimentaria, así como para evitar prácticas fraudulentas (Ababouch *et al.*, 2005; Freitas *et al.*, 2020). Además, es importante para detectar escapes de organismos de cultivo al medio silvestre, lo cual puede afectar la composición genética de las poblaciones silvestres (Utter, 2003; Danancher y Garcia-Vazquez, 2011; Glover *et al.*, 2013).

Los avances en la genómica han permitido, a través de las técnicas de secuenciación masiva, el desarrollo de paneles de marcadores moleculares distribuidos a lo largo del genoma, en especial marcadores tipo Polimorfismo de una sola Base (SNPs por sus siglas en inglés), los cuales tienen el potencial de ser utilizados para incrementar la precisión en la selección genética de rasgos de interés comercial, incluso en caracteres difíciles de seleccionar directamente, como la resistencia a enfermedades (Castillo-Juárez *et al.*, 2015; Zenger *et al.*, 2019). La obtención de dichos paneles de SNPs en especies de cultivo se han empezado a implementar en estudios de genética de poblaciones e identificación de stocks, análisis de parentesco, así como Selección Genómica (GS) y Estudios de Asociación a lo largo del Genoma (GWAS por sus siglas en inglés). Estos últimos permiten la implementación de estrategias adecuadas de selección basadas en información genómica (Palaikostas y Houston, 2017; Zenger *et al.*, 2019).

Los marcadores moleculares tipo SNP tiene diversas ventajas, en comparación con otro tipo de marcadores, entre otras, presentan menor error de genotipificación, menor costo en el análisis de un gran número de muestras debido a que pueden ser analizadas simultáneamente (Vignal *et al.*, 2002; Hayes *et al.*, 2005).

En este trabajo, se realizó la caracterización genómica de poblaciones de cultivo y silvestres de *P. vannamei* en México con paneles de SNPs de baja densidad, así como su validación para análisis de parentesco e identificación de origen poblacional.

## 2. ANTECEDENTES

### 2.1 Marcadores moleculares y técnicas de obtención de genotipos

Los marcadores moleculares son fundamentales en estudios genéticos de diversidad, endogamia, genética de poblaciones, análisis de parentesco, identificación de especies e híbridos, así como mapas de ligamiento, estudios de asociación a lo largo del genoma (GWAS), selección genómica (GS), entre otros. Con el tiempo, se han desarrollado y utilizado diferentes tipos de marcadores moleculares; sus principios y aplicaciones en especies acuícolas se ha discutido ampliamente (e.g. Liu y Cordes, 2004; Liu, 2007; Chauhan y Rajiv, 2010).

En *P. vannamei*, los marcadores tipo microsatélites han sido tradicionalmente los más populares debido a su alto polimorfismo, y se han desarrollado y aplicado en los últimos 25 años (e.g. Wolfus *et al.*, 1997; Cruz *et al.*, 2002; Meehan *et al.*, 2003; Zhi-Ying *et al.*, 2006; Garcia y Alcivar-Warren, 2007; Santos *et al.*, 2012; Perez-Enriquez *et al.*, 2018a), centrándose en la caracterización de poblaciones silvestres y de cultivo, diversidad genética, parentesco, estimación de parámetros genéticos, entre otros. Sin embargo, debido a la presencia de alelos nulos derivado de altas tasas de mutación (Ellegren, 2000), dificultad para estandarizar la obtención de genotipos entre distintos laboratorios, entre otras desventajas, estos han sido sustituidos en los últimos años por polimorfismos de base sencilla (SNPs).

Los SNPs son marcadores genéticos que consisten en la variación de un nucleótido (A, C, G o T) en una posición específica del genoma; ofrecen varias ventajas sobre los microsatélites y otros marcadores genéticos, como la abundancia y distribución a lo largo del genoma (Collins *et al.*, 1998; Rafalski, 2002), bajos errores en obtención de genotipos, con menor costo y tiempo de análisis de muestras (Vignal *et al.*, 2002; Hayes *et al.*, 2005), entre otras. A pesar de ser menos polimórficos (los SNPs son en su mayoría bialélicos), la identificación y genotipificación de un gran número de SNPs a lo largo del genoma, en gran cantidad de individuos, se ha vuelto más accesible. Avances en métodos de secuenciación masiva han llevado a la popularización de los SNPs, p.ej. la secuenciación de ADN asociada a sitios de restricción (RAD-seq) (Baird *et al.*, 2008) se ha aplicado ampliamente en la acuicultura (Robledo *et al.*, 2018). En los últimos diez años han surgido diferentes variaciones de RAD-seq, como 2b-RAD (Wang *et al.*, 2012), SLAF-seq (Sun *et al.*, 2013),

dd-RAD (Peterson *et al.*, 2012) y ezRAD (Toonen *et al.*, 2013), todos ellos basados en el uso de enzimas de restricción, lo que reduce la representación del genoma y, por tanto, reduce el coste de la secuenciación. Otra ventaja de las tecnologías tipo RAD es que no requieren un ensamblaje del genoma a nivel cromosómico como referencia.

En el camarón blanco *P. vannamei*, 2b-RAD (Wang *et al.*, 2017b; Wang *et al.*, 2019a; b; Dai *et al.*, 2020; Lyu *et al.*, 2021; Chen *et al.*, 2022), SLAF-seq (Yu *et al.*, 2015; Yu *et al.*, 2019; Wang *et al.*, 2017a; Peng *et al.*, 2020) y NextRAD (Perez-Enriquez *et al.*, 2018c) han sido las técnicas más aplicadas. Estos métodos permitieron la caracterización genética, la construcción de mapas de ligamiento, GWAS y GS.

Una vez que se ha identificado una gran cantidad de SNPs (paneles de alta densidad con miles de marcadores), es posible desarrollar arreglos para su genotipificación, los cuales ya están disponibles para varias especies de cultivo, por ejemplo: salmón del Atlántico (Houston *et al.*, 2014; Yáñez *et al.*, 2016), lubina y dorada (Peñaloza *et al.*, 2021), tilapia (Peñaloza *et al.*, 2020), ostra del Pacífico (Qi *et al.*, 2017), entre otras.

Para *P. vannamei*, hay dos paneles de SNPs de alta densidad disponibles comercialmente *Illumina Infinium Shrimp LD-24 BeadChip* con ~6,4k SNP y el *AQUArray HD vannamei* con 50k SNP; este último fue desarrollado por Neogen y el Centro de Tecnologías Acuícolas (Center of Aquaculture Technologies, CAT; <https://aquatechcenter.com/>). Estas tecnologías se están utilizando con éxito para la identificación de QTL (Jones *et al.*, 2020) y GWAS (Medrano-Mendoza *et al.*, 2023), pero su aplicación en operaciones comerciales aún está limitada por el costo por individuo.

Para aplicaciones que solo requieren de decenas a cientos de marcadores genéticos, como análisis de pedigrí, han surgido otras técnicas de genotipificación para satisfacer esta demanda: *AQUArray LD vannamei* (192 SNPs; CAT, <https://aquatechcenter.com/>), Fluidigm (<https://www.fluidigm.com/>), genotipificación en miles mediante secuenciación (GT-seq) (Campbell *et al.*, 2015), Secuenciación de amplicones dirigidos de PCR multiplex (MTA-seq) (Onda *et al.*, 2018), Secuenciación de amplicones altamente multiplexados (HiMAP) (Dupuis *et al.*, 2018),

ó Ion Ampliseq de Thermo Fisher. Estos se basan principalmente en la secuenciación de amplicones de PCR multiplex de SNPs previamente identificados.

Otras alternativas se basan en amplicones de PCR marcados con fluoróforos, como KASP (Biosearch Technologies; <https://www.biosearchtech.com/>), TaqMan (Roche Molecular Systems, Inc.) y Curvas de disociación de alta resolución (HRM), o aquellos basados en espectrometría de masas como MassARRAY System (Sequenom; San Diego, California), que pueden ser más apropiados para genotipificar pocos marcadores. Por ejemplo, en *P. vannamei* HRM se ha utilizado para la genotipificación de un SNP de identificación sexual (Perez-Enriquez *et al.*, 2020), mientras que la plataforma Sequenom MassARRAY iPLEX Gold se ha utilizado para genotipificar 49 SNPs presumiblemente asociados con la tolerancia al amoníaco (Lu *et al.*, 2018).

Al momento de seleccionar un panel SNPs y el método de genotipificación en *P. vannamei*, se deben tener en cuenta diferentes factores, como el objetivo del estudio y los recursos financieros disponibles. En el cultivo de camarón blanco, con un valor relativamente bajo por animal [~55 USD por un reproductor de *P. vannamei* certificado como libre de patógenos específicos (SPF) (Van der Pijl, 2020)], se requiere una plataforma de genotipificación de SNPs rentable para análisis genéticos de rutina en programas de reproducción, especialmente para pequeños y medianos productores. Hasta el momento, las plataformas con el menor costo por individuo actualmente disponibles para la especie son *AQUAarray HD* y *AQUAarray LD*, para paneles SNP de alta y baja densidad, respectivamente. El desarrollo de paneles de baja densidad basados en GT-seq (Campbell *et al.*, 2015), y 2bRAD utilizando adaptadores selectivos (Barbanti *et al.*, 2020) pueden ser una buena alternativa para disminuir los costos de obtención de genotipos por individuo, aunado a esto, evitaría los problemas de logística del envío de muestras al extranjero.

## **2.2 Caracterización genética de poblaciones de cultivo y silvestres**

El camarón blanco *P. vannamei* se distribuye a lo largo de la costa oriental del Océano Pacífico, desde México hasta Perú (FAO, 2006). A lo largo de su distribución se ha reportado una estructura genética poblacional en distintas subpoblaciones utilizando ADN mitocondrial (Valles-Jimenez *et al.*, 2006), microsatélites (Valles-Jimenez *et al.*, 2005) y SNPs (Perez-Enriquez *et al.*, 2024b)

explicada por barreras oceanográficas, sin embargo, entre el noroeste y el suroeste del Pacífico podría deberse al aislamiento por distancia (Perez-Enriquez *et al.*, 2024b). En escala regional, en el Golfo de California no se ha observado diferenciación genética entre poblaciones (Perez-Enriquez *et al.*, 2018b).

*P. vannamei* es cultivado a nivel mundial en más de 30 países, siendo los 10 principales productores China, Indonesia, India, Ecuador, Vietnam, Tailandia, México, Brasil, Arabia Saudita e Iran (FAO, 2020). En México, la caracterización genética de líneas de cultivo ha sido previamente analizada en la que se reportan valores medios de variabilidad genética en seis cultivos con seis loci microsatélites, además de no encontrar signos de pérdida de variabilidad a lo largo del tiempo (comparación entre 2000-2002 y 2007); por otra parte, encontraron diferencias significativas entre las líneas de cultivo y una población silvestre, así como entre la mayoría de las poblaciones de cultivo (Perez-Enriquez *et al.*, 2009). En otro estudio, con marcadores mitocondriales, analizaron muestras de 5 laboratorios y una muestra silvestre; detectan dos haplotipos dominantes en todas las poblaciones, incluyendo la muestra silvestre, y sugieren la hipótesis de un origen común (Mendoza-Cano *et al.*, 2013).

En 2013, 2014 y 2015 ocurrieron eventos de importación en México de líneas de cultivo de distintos orígenes, esto como respuesta de los productores acuícolas para mejorar la resistencia de sus lotes de cultivo (Perez-Enriquez *et al.*, 2018c). Perez-Enriquez *et al.* (2018c) con un panel de 2,619 SNPs, detectaron de 3 a 5 grupos genéticos con distintos orígenes, estos resultados tienen el potencial para desarrollar un panel de marcadores para la identificación de origen y trazabilidad de los organismos.

En programas de mejoramiento genético y reproducción selectiva, un elemento esencial es la caracterización de la diversidad y composición genética de lotes de cultivo para conocer su estado genético, tomar decisiones para incrementar la variabilidad y/o evitar su pérdida, así como la posible trazabilidad de los organismos.

### 2.3 Trazabilidad y asignación poblacional

La trazabilidad cobra importancia en la acuicultura con la necesidad de monitorear los productos acuícolas a lo largo de la cadena productiva, así como con la autenticación de origen (p.ej. si son de cultivo o silvestres), o comprobar el etiquetado de especie, lo cual es clave para la seguridad alimentaria, así como para evitar prácticas fraudulentas (Ababouch *et al.*, 2005; Freitas *et al.*, 2020). Los marcadores moleculares tienen el potencial de ser implementados para este fin como una herramienta de trazabilidad tanto en la acuicultura como en pesquerías (Hastein *et al.*, 2001; Ogden, 2008; Martinsohn *et al.*, 2019).

El uso de marcadores moleculares tipo SNPs ha sido evaluado para su uso con fines de trazabilidad, principalmente en pesquerías para identificar la población/región de origen de los organismos, como en la pesquería de salmón real *Oncorhynchus tshawytscha* en la costa de Estados Unidos (Clemento *et al.*, 2014), en el atún aleta azul *Thunnus thynnus* (Rodríguez-Ezpeleta *et al.*, 2019), y en cuatro especies pesqueras en Europa (bacalao, arenque, lenguado y merluza) (Nielsen *et al.*, 2012), este último estudio utiliza SNPs asociadas a genes, con los que logran identificar la población de origen de los organismos con alta precisión (93-100%). En organismos acuícolas, en el cultivo de salmón del Atlántico en Noruega, se evaluaron varios esquemas de trazabilidad con distintos marcadores moleculares (microsatélites y SNPs), en donde los SNPs resultaron una mejor estrategia en cuanto a costos se refiere (Hayes *et al.*, 2005); en esta misma especie, otro estudio con un panel reducido de 94SNPs, evaluó la asignación de individuos a una de tres poblaciones de cultivo (dos de origen noruego y una de origen escocés) (Holman *et al.*, 2017). Hasta la fecha, en *P. vannamei*, no se han realizado estudios para evaluar la identificación de origen poblacional de organismos, tanto de cultivo como silvestre, con fines de trazabilidad.

Por otra parte, la importancia de la trazabilidad radica en la detección de escapes de individuos de cultivo al medio ambiente, lo cual puede ocasionar un impacto negativo en las poblaciones silvestres afectando su composición genética (Utter, 2003; Danancher y Garcia-Vazquez, 2011; Glover *et al.*, 2013). La presencia y/o introgresión de organismos de cultivo se ha detectado con el uso de microsatélites en distintas especies como trucha arcoíris *Oncorhynchus mykiss* (Glover, 2008, 2010), salmón del Atlántico *Salmo salar* (Glover *et al.*, 2008; Glover *et al.*, 2009; Zhang *et*

*al.*, 2013), lubina europea *Dicentrarchus labrax* (Brown *et al.*, 2015); así como en camarón blanco *P. vannamei* (Perez-Enriquez *et al.*, 2018a), en donde la presencia de organismos de cultivo en una muestra silvestre es aparentemente baja (7.1%) y no representa un gran impacto a la fecha del estudio; recientemente, el uso de SNPs ha sido evaluado con este fin en rodaballo *Scophthalmus maximus* (Prado *et al.*, 2018), lenguado y bacalao (Bylemans *et al.*, 2016); así como en salmón del Atlántico (Glover *et al.*, 2013; Karlsson *et al.*, 2016) en donde se detectaron zonas con alto grado de introgresión.

## **2.4 Análisis de pedigrí en la acuicultura**

El seguimiento del pedigrí en la acuicultura permite definir esquemas adecuados de entrecruzamiento, minimizar la endogamia y, por lo tanto, evitar la depresión por endogamia (Gjerde y Rye, 1998; Gjedrem y Baranski, 2009). Sin embargo, el marcaje físico en etapas tempranas en muchas especies no es posible por su tamaño, como en el caso de *P. vannamei*, por lo que, sería necesario mantener una gran cantidad de familias en tanques separados para lo que se necesitaría mayor infraestructura lo que podría no ser viable para los productores. La alternativa que ha sido aplicada en varias especies acuícolas es el uso de marcadores moleculares en análisis de parentesco (Vandeputte y Haffray, 2014). Desde mediados de los 90s, la asignación de parentesco con el uso de microsatélites se introdujo como herramienta en programas de acuicultura (Herbinger *et al.*, 1995; Estoup *et al.*, 1998) y han sido comúnmente usados en distintas especies, incluidos los penaeidos (Jerry *et al.*, 2004; Jerry *et al.*, 2006; Dong *et al.*, 2006). El objetivo final de un panel de marcadores para análisis de parentesco es la asignación del 100% de la progenie con un alto grado de certidumbre y al menor costo posible. Los avances en las tecnologías de secuenciación masiva y obtención de genotipos han permitido el uso de marcadores alternativos, como los SNPs, los cuales han demostrado una mayor eficacia y robustez en los análisis de parentesco, así como menos errores de genotipificación, tanto en camarón blanco (Perez-Enriquez y Max-Aguilar, 2016), como en otras especies acuícolas (Sellars *et al.*, 2014; Liu *et al.*, 2017; Zhao *et al.*, 2018). El desarrollo y validación de paneles de SNPs en análisis de pedigrí en la acuicultura han sido foco de investigación en años recientes (Liu *et al.*, 2016; Holman *et al.*, 2017; Harney *et al.*, 2018; Premachandra *et al.*, 2019).



Como se mencionó en la sección 2.1, para *P.vannamei* existe un panel de SNPs comercial disponible (AQUAarray LD; CAT), sin embargo, los costos de genotipificación de este panel (~\$12 USD por muestra), aunado a la problemática de envío de muestras a USA (p. ej. la retención de las muestras en la aduana) hace que este no sea la mejor opción para análisis de rutina comercial en el cultivo de camarón, especialmente en pequeños/medianos productores, por lo que, se hace necesario el desarrollo y validación de un panel de SNPs que sea más asequible para la industria camaronícola de México.

### 3. JUSTIFICACIÓN

El camarón blanco *P. vannamei* es una de las especies acuícolas más importantes a nivel mundial y la más importante en México por su valor comercial, la producción acuícola puede ser más eficiente a través de la implementación de programas de selección y mejoramiento genético, y estos requieren la implementación de herramientas moleculares para el monitoreo del pedigrí, caracterización e identificación genética de líneas de cultivo, así como su trazabilidad, en especial en líneas con características genéticas de interés comercial.

La relevancia económica del presente trabajo radica en esta necesidad de optimizar la producción de camarón con la implementación de programas de selección genética, los cuales han demostrado que mejoran la eficiencia de los sistemas productivos, comparado con sistemas tradicionales. Por ejemplo, con programas de selección enfocados a resistencia a enfermedades como parte de una estrategia para prevenir y evitar pérdidas masivas por la presencia de patógenos (las pérdidas de producción en 2013 respecto al año anterior equivaldrían a 1,900-2,600 millones de pesos). Además, los programas de manejo de reproductores y mejoramiento genético requieren de la implementación de estrategias de monitoreo de la diversidad y del pedigrí, para evitar la pérdida de variabilidad y la endogamia a través de las generaciones de cultivo, lo cual puede llevar a la depresión por endogamia y a un bajo desempeño en la producción en el corto plazo. Este trabajo implementa el uso de paneles de SNPs de baja densidad para estos análisis, que además sean más asequibles para la industria, evitando el envío de muestras al extranjero para su análisis.

Por otro lado, como parte de la importancia ambiental, el implementar estrategias de trazabilidad en la especie permite detectar prácticas fraudulentas, como la detección de pesca ilegal durante las épocas de veda, esto a través del análisis de los perfiles genéticos de los organismos y compararlo con los perfiles de su presunta población de origen; además, las herramientas moleculares desarrolladas en este trabajo pueden ser implementadas para evaluar el grado de introgresión de organismos de cultivo al medio ambiente, lo cual puede impactar negativamente a las poblaciones silvestres.

Como relevancia social, la camaronicultura involucra el empleo directo e indirecto de aproximadamente 42,000 personas en México; el crecimiento de esta industria generaría mayor cantidad de empleos, así como ayudaría a cubrir la creciente demanda de proteína animal para consumo humano.

El aporte científico y tecnológico del trabajo está enfocado en la caracterización genética de las poblaciones de cultivo en México, así como en el desarrollo de paneles de SNPs y la estandarización de técnicas de obtención de genotipos (como GT-seq y 2bRAD), que en este estudio permitieron conocer el estatus genético de las poblaciones en México, pero que están disponibles para otros fines de investigación, como para GWAS (estudios de asociación a lo largo del genoma) que permitirían entender la arquitectura genética de caracteres de interés.

#### **4. HIPÓTESIS**

La caracterización genómica de poblaciones de cultivo de camarón blanco *Penaeus vannamei* mediante Polimorfismos de Base Sencilla (SNPs) de baja densidad, permitirá la estimación de los parámetros genéticos requeridos para la trazabilidad genética y seguimiento de parentesco.

## 5. OBJETIVOS

### 5.1 Objetivo general

Definir estrategias para la trazabilidad y análisis de parentesco en camarón blanco *Penaeus vannamei*

### 5.2 Objetivos particulares

- Desarrollar paneles de marcadores moleculares tipo SNPs en camarón blanco *Penaeus vannamei*
- Caracterizar genómicamente poblaciones de cultivo de *Penaeus vannamei* provenientes de distintos orígenes con un panel de baja densidad de SNPs y validarlo para análisis de trazabilidad.
- Validar un panel de baja densidad de SNPs para análisis de parentesco

## 6. MATERIAL Y MÉTODOS

### 6.1 Desarrollo de paneles de marcadores moleculares tipo SNP en camarón blanco *P. vannamei*

Se siguieron dos estrategias para la obtención de SNPs: 1) se desarrolló un panel de SNPs *de novo* mediante la técnica de secuenciación Asociada a Sitios de Restricción con enzimas IIB (2bRAD; sección 6.1.1); y 2) se desarrolló un panel de SNPs de baja densidad a partir de una base de datos de secuenciación pública y se configuró teniendo en consideración la genotipificación en miles por secuenciación (GT-seq) y curvas de disociación de alta resolución (HRM) (sección 6.1.2).

#### 6.1.1 Obtención de panel de SNPs *de novo* con 2bRAD

El método de 2bRAD (Wang *et al.*, 2012) consiste en la digestión del ADN con una enzima de restricción tipo IIB que identifica una secuencia específica y corta hacia ambos lados del sitio de reconocimiento, con lo cual se obtienen fragmentos del mismo tamaño con extremos cohesivos (nucleótidos en rojo en Fig. 2). Posteriormente, se adicionan adaptadores a dichos extremos, los cuales pueden tener bases selectivas y con ello, disminuir el número de fragmentos y SNPs obtenidos. A los fragmentos también se adicionan códigos de barra para la identificación de cada individuo analizado, los cuales son secuenciados para la obtención de los genotipos.

Para la obtención de un panel SNPs *de novo* se realizó un análisis *in silico* para determinar el número de fragmentos esperados para cuatro enzimas de restricción tipo IIB, seleccionar la más adecuada y estimar la proporción fragmentos con variantes tipo SNP (sección 6.1.1.1); posteriormente se estandarizó la técnica de 2bRAD y se identificó el número de SNPs obtenido para el panel (sección 6.1.1.2).

##### 6.1.1.1 Evaluación *in silico* de enzimas de restricción tipo IIB

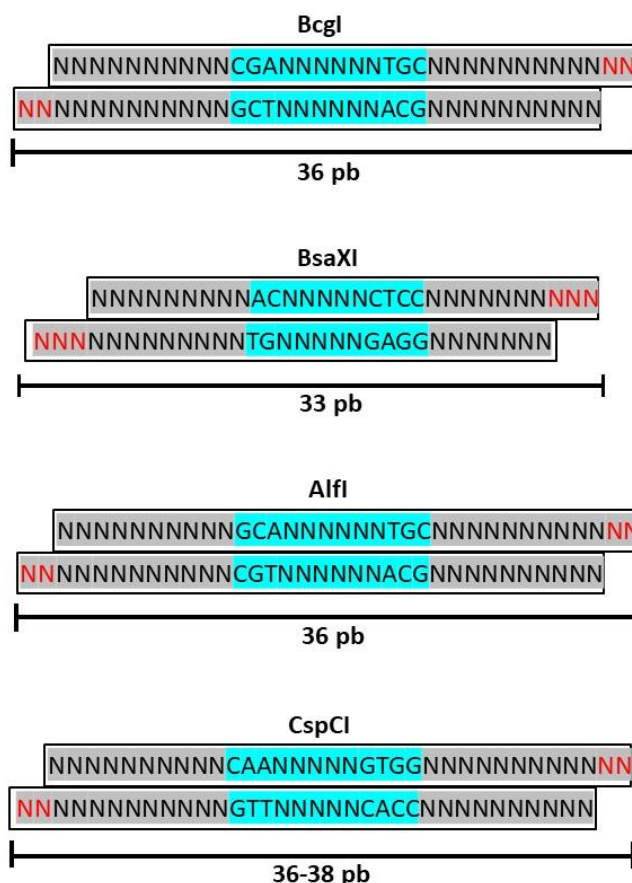
El genoma de referencia de *P. vannamei* publicado por Zhang *et al.* (2019) (GenBank no. Acceso ASM378908v1) fue utilizado para identificar los sitios de reconocimiento de cuatro enzimas de restricción tipo IIB (Tabla 1). El genoma se descargó de la base de datos de GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) al servidor del Laboratorio de Genómica y Bioinformática del CIBNOR. Se contabilizó el número de sitios de reconocimiento de cada enzima

en dirección sentido y anti-sentido en Linux. Los comandos utilizados se encuentran en el Anexo A.

**Tabla 1.** Secuencias de reconocimiento y corte de enzimas de restricción tipo IIB y tamaño del fragmento obtenido.

Enzima	Secuencia de reconocimiento y corte	Tamaño del fragmento
Bcgl	5' <sub>10</sub> (N)CGA(N) <sub>6</sub> TGC(N) <sub>12</sub> 3' 3' <sub>12</sub> (N)GCT(N) <sub>6</sub> ACG(N) <sub>10</sub> 5'	36
BsaXI	5' <sub>9</sub> (N)AC(N) <sub>5</sub> CTCC(N) <sub>10</sub> 3' 3' <sub>12</sub> (N)TG(N) <sub>5</sub> GAGG(N) <sub>7</sub> 5'	33
AlfI	5' <sub>10</sub> (N)GCA(N) <sub>6</sub> TGC(N) <sub>12</sub> 3' 3' <sub>12</sub> (N)CGT(N) <sub>6</sub> ACG(N) <sub>10</sub> 5'	36
CspCI	5' <sub>10-11</sub> (N)CAA(N) <sub>5</sub> GTGG(N) <sub>12-13</sub> 3' 3' <sub>12-13</sub> (N)GTT(N) <sub>5</sub> CAGG(N) <sub>10-11</sub> 5'	36-38

El número de fragmentos obtenido y por lo tanto de SNPs, puede ser reducido al incluir adaptadores con bases selectivas, que se unirán por complementariedad a los extremos cohesivos. Parte del objetivo de este panel es que pueda ser de baja densidad (decenas de SNPs) para ser utilizado en análisis de parentesco, por lo tanto, se estimó el número de fragmentos remanentes si se incluyen de 1-6 bases selectivas, según el número posible de bases selectivas a incluir por enzima (Fig. 2). Los comandos utilizados se incluyen en el Anexo A.



**Figura 2.** Sitios de reconocimiento y corte de enzimas tipo IIB. En azul, el sitio de reconocimiento de cada enzima. En rojo, las bases nucleotídicas que corresponden a los extremos cohesivos y que, en los adaptadores, pueden ser substituidas por bases selectivas.

Posteriormente, se estimó *in silico* la proporción de sitios variables respecto al número de sitios de reconocimiento con la enzima de restricción seleccionada. Para ello, se utilizó la base de datos de la secuenciación de Perez-Enriquez *et al.* (2018c) (NCBI-SRA Bioproject PRJNA492152). En Linux, se seleccionaron los fragmentos con el sitio de reconocimiento de la enzima seleccionada con el comando `zgrep` y se alinearon al genoma de referencia de *P. vannamei* (Zhang *et al.*, 2019) utilizando Burrows-Wheeler Aligner (`bwa`) (Li y Durbin, 2009) y `samtools` (Li *et al.*, 2009). La obtención de SNPs se realizó con `samtools` y `bcftools` (Li *et al.*, 2009). Posteriormente, se hizo el filtrado de los SNPs obtenidos con `vcftools` (Danecek *et al.*, 2011) con los siguientes criterios: MAF (Frecuencia Mínima de Alelo Menor) de 0.016, `minQ` (valor de calidad mínima) de 30, proporción de individuos en los que se encuentra cada SNP de 1.0, remoción de *indels* (polimorfismos de inserción/delección), máximo y mínimo de alelos por SNP de 2. De los fragmentos que presentaban SNPs, se revisaron al azar 50 de ellos con el programa Integrative Genomics Viewer (IGV)



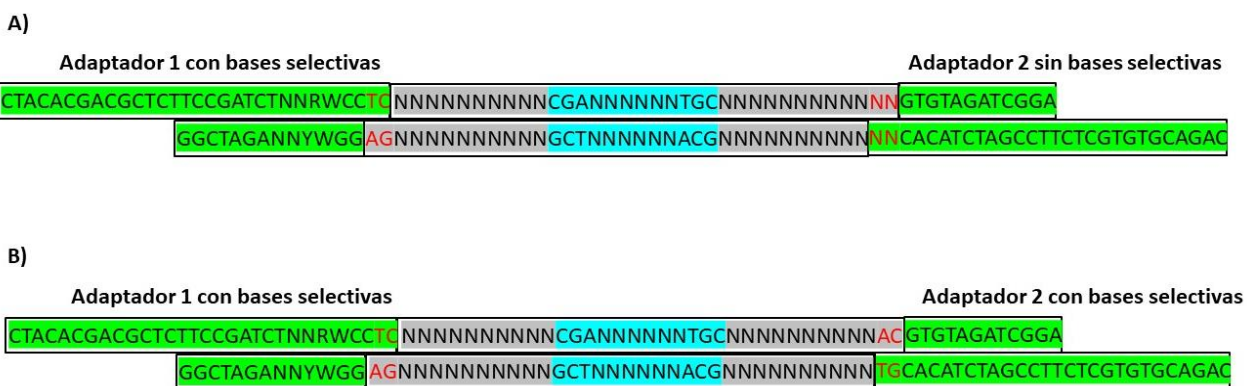
(Robinson *et al.*, 2011) para buscar en qué proporción de ellos, los SNPs se encontraban dentro del fragmento que se obtendría con la enzima de digestión seleccionada.

### 6.1.1.2 Estandarización de 2bRAD para obtención de panel de SNPs

La técnica de 2bRAD se estandarizó con un número reducido de muestras para verificar el número de marcadores obtenidos con la enzima de restricción elegida en el análisis *in silico* de la sección 6.1.1.1.

Se tomaron muestras de tejido de ocho camarones: cuatro provenientes de cultivo y cuatro de origen silvestre. Se realizó la extracción de ADN con el kit DNeasy blood and tissue de Qiagen siguiendo las instrucciones del fabricante. El ADN se cuantificó en un Nanodrop 2000 y se verificó que la pureza fuera adecuada (proporción 260/280 de 1.8-2.2 y proporción 260/230  $\geq$ 1.8). Se prepararon alícuotas de ADN a 50 ng/ $\mu$ l y se cuantificaron nuevamente en un Qubit con el kit dsDNA HS siguiendo las instrucciones del fabricante. Considerando las lecturas por Qubit, se preparó el ADN a  $\sim$ 15 ng/ $\mu$ l.

De las ocho muestras de ADN se prepararon dos pools de ADN con cuatro individuos cada uno: 1) con las cuatro muestras de cultivo y 2) con las cuatro muestras silvestres. Con cada pool de ADN se prepararon dos librerías distintas: 1) utilizando el adaptador 1 con dos bases selectivas (TC) y el adaptador 2 sin bases selectivas (Fig. 3A); 2) utilizando ambos adaptadores con bases selectivas (Fig. 3B). Por lo que se prepararon cuatro librerías 2bRAD en total.



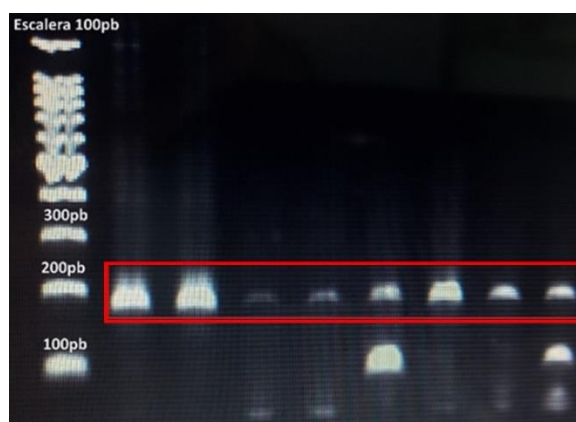
**Figura 3.** Fragmentos de ADN después de la ligación de adaptadores. Adaptadores de doble cadena en color verde. A) Ligación de adaptador 1 con bases selectivas TC y adaptador 2 sin bases selectivas. B) Ligación de adaptadores 1 y 2 con bases selectivas.

Para la preparación de las librerías se siguió el protocolo de Faust *et al.* (2018) modificado. La preparación consistió en 5 etapas: 1) Digestión con enzima Bcgl, 2) Ligación de adaptadores, 3) Amplificación y adición de códigos de barra, 4) Purificación de fragmentos y 5) Secuenciación. La secuencia de los adaptadores y oligos para la preparación de librerías se encuentran en el Anexo B.

1) *Digestión*. Se realizó la digestión con la enzima Bcgl y 60 ng de ADN de cada muestra (pool cultivo y pool silvestre) por duplicado. La digestión se realizó en 6  $\mu$ l que contenían 0.6  $\mu$ l de buffer de enzima (10 $\times$ ), 0.9  $\mu$ l de agua libre de nucleasas, 0.5  $\mu$ l de enzima Bcgl (2 U/ $\mu$ l) y 4  $\mu$ l de ADN (15 ng/ $\mu$ l), se incubaron a 37°C por 1 hora en un termociclador Bio-rad C1000 Touch; posterior a la incubación, se inactivó la enzima a 65°C por 10 min.

2) *Ligación de adaptadores*. Como se mencionó previamente, se prepararon dos librerías con la muestra de cultivo y dos con la muestra silvestre utilizando adaptadores de doble cadena: 1) con el adaptador 1 con bases selectivas y el adaptador 2 sin bases selectivas, y 2) con ambos adaptadores con bases selectivas (Fig. 3). La ligación se realizó en 26  $\mu$ l de volumen final que contenían 2  $\mu$ l de buffer de enzima T4 ligasa (10 $\times$ ), 0.5  $\mu$ l de adaptador 1, 0.5  $\mu$ l de adaptador 2, 1  $\mu$ l de T4 DNA ligasa (400 U/  $\mu$ l), 16  $\mu$ l de agua libre de nucleasas y 6  $\mu$ l de la digestión realizada en el paso anterior. Se incubaron a temperatura ambiente durante toda la noche y posteriormente, se inactivó la ligasa a 65°C por 20 min. Se preservaron a 4°C.

3) *Amplificación y adición de códigos de barra*. Se realizó en una reacción con volumen final de 50  $\mu$ l que contenía: 10  $\mu$ l de la ligación, 3  $\mu$ l de oligo ILL-BC23-30 (2  $\mu$ M; uno distinto para identificar cada librería), 3  $\mu$ l de oligo Truseq-UN01 (2  $\mu$ M), 1  $\mu$ l Taq polimerasa Promega (5 U/ $\mu$ l), 10  $\mu$ l de buffer para Taq polimerasa (5 $\times$ ), 4  $\mu$ l de MgCl<sub>2</sub> (25 mM), 1  $\mu$ l de dNTPs mix (10 mM), 1  $\mu$ l de oligo P5 (10  $\mu$ M), 1  $\mu$ l de oligo P7 (10  $\mu$ M) y 16  $\mu$ l de agua MilliQ. La amplificación se realizó en un termociclador Bio-rad C1000 Touch con el siguiente protocolo: 70°C por 30 s; 95°C por 2 min; 15 ciclos de 95°C 20 s, 65°C 3 min, 72°C 30 s. Se visualizaron los fragmentos esperados de ~180 pb en un gel de agarosa al 2%. En la Fig. 4 se observan los fragmentos esperados señalados en el rectángulo rojo.



**Figura 4.** Visualización de librerías 2bRAD en gel de agarosa al 2% previo a la purificación de fragmentos. El fragmento esperado de 180 pb se señala en el rectángulo rojo.

4) *Purificación de fragmentos.* Se realizó una purificación inicial de 40  $\mu$ l de cada librería con 1.3 $\times$  de perlas magnéticas AMPure XP beads (Beckman Coulter). Se preparó una librería final mezclando 30  $\mu$ l de cada librería individual y se realizó una segunda purificación del fragmento de  $\sim$ 180 pb con PippinPrep en un casete de 2% agarosa Dye-Free para 100-600 pb siguiendo las instrucciones del fabricante.

5) *Secuenciación.* La librería final se cuantificó por duplicado en un Qubit con el kit dsDNA HS y se preparó a 2 nM. Previo a la secuenciación, la librería fue desnaturalizada y diluida a 10 pM. La secuenciación se realizó en un Miseq Illumina en el Laboratorio de Genómica y Bioinformática del CIBNOR, con una celda v3 de 150 ciclos y 25 millones de lecturas, y 20% de PhiX.

Se descargaron los archivos Fastq generados de la secuenciación y se analizó la calidad de las lecturas con FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Se eliminaron las secuencias de baja calidad y los adaptadores con el programa Trimmomatic (Bolger *et al.*, 2014). Las secuencias se alinearon al genoma de referencia de *P. vannamei* (Zhang *et al.*, 2019) con Burrows-Wheeler Aligner (bwa) (Li y Durbin, 2009) y samtools (Li *et al.*, 2009). La obtención de SNPs se realizó con samtools, bcftools (Li *et al.*, 2009) y vcftools (Danecek *et al.*, 2011).

## 6.1.2 Obtención de panel SNPs utilizando datos públicos de secuenciación

Se desarrolló un panel de baja densidad de marcadores moleculares tipo SNP utilizando datos públicos de secuenciación. Esto requirió de análisis bioinformáticos para la obtención, filtrado y selección de SNPs (sección 6.1.2.1); se diseñaron oligos para la amplificación de los SNPs seleccionados y se realizó un análisis *in silico* para evaluar la posibilidad de amplificar todos los SNPs en una misma reacción multiplex (sección 6.1.2.2). Parte del proceso de selección del panel de SNPs incluyó su validación *in silico* para ser utilizado en análisis de asignación poblacional y estructura genética, así como para análisis de parentesco, estos se incluyeron en la sección 6.2.1 y 6.3.1 respectivamente.

### 6.1.2.1 Análisis bioinformático para obtención de SNPs

Para la obtención de SNPs se utilizaron los datos de la secuenciación del proyecto NCBI-SRA PRJNA492152 (Perez-Enriquez *et al.*, 2018c) del Genebank. Las secuencias fueron revisadas con FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) para verificar su calidad y se recortaron las lecturas de baja calidad con Trimmomatic (Bolger *et al.*, 2014). Las secuencias fueron alineadas al genoma de camarón *P. vannamei* reportado en Zhang *et al.* (2019) utilizando el programa Burrows-Wheeler Aligner (bwa) (Li y Durbin, 2009) y samtools (Li *et al.*, 2009). La obtención de SNPs se realizó con samtools y bcftools (Li *et al.*, 2009).

Se seleccionaron SNPs para dos objetivos distintos: 1) para estructura genética y asignación poblacional, y 2) para análisis de parentesco. Por lo que, a partir del siguiente paso, se desarrollaron dos bases de datos distintas.

1) *SNPs para estructura genética y asignación poblacional.* Se filtraron los SNPs con vcftools (Danecek *et al.*, 2011) con los siguientes criterios: MAF (Frecuencia Mínima de Alelo Menor) de 0.016, minQ (valor de calidad mínima) de 30, proporción de individuos en los que se encuentra el SNP de 1.0, máximo y mínimo de alelos por SNP de 2, y remoción de *indels* (polimorfismos de inserción/delección). Después del filtrado, se seleccionaron los SNPs que tuvieran un mínimo de 30 bases a cada lado libre de variantes, con la finalidad de tener una región conservada donde se pudieran diseñar oligos para su amplificación por PCR. Con los SNPs resultantes, se realizaron análisis de diferenciación genética de *Fst* de forma pareada con vcftools (Danecek *et al.*, 2011).

Para este se consideraron los cinco grupos genéticos de cultivo en México reportados en Perez-Enriquez *et al.* (2018c) lo que resultó en 10 comparaciones pareadas. Se seleccionaron los SNPs con los valores de diferenciación más altos. Un filtro adicional fue la eliminación de los SNPs con genotipos A/T, dado que, para evaluar la obtención de genotipos mediante curvas de disociación (HRM), la identificación de este par de alelos es difícil con esta técnica.

2) *SNPs para análisis de parentesco*. Se filtraron los SNPs con vcftools (Danecek *et al.*, 2011) con los siguientes criterios: MAF de 0.2, minQ de 30, proporción de individuos en los que se encuentra el SNP de 0.7, máximo y mínimo de 2 alelos por SNP, y remoción de *indels*. Después del filtrado, se seleccionaron los SNPs que tuvieran un mínimo de 30 bases a cada lado libre de variantes, con la finalidad de tener una región conservada donde se pudieran diseñar oligos para su amplificación por PCR. Además, se dio prioridad a seleccionar SNPs que dentro del fragmento de 150 bases tuvieran 1 o 2 SNPs, con el objetivo de facilitar el diseño de oligos.

Con el panel final de SNPs se realizó un análisis de desequilibrio de ligamiento pareado utilizando Arlequin 3.5.2.2 (Excoffier y Lischer, 2010), para verificar que no existía asociación entre los loci con pruebas de probabilidad (likelihood-ratio) con 10,000 permutaciones y el número de condiciones iniciales aleatorias para el algoritmo EM (*Expectation-Maximization*) de 3. Para esta prueba sólo se utilizaron individuos de un mismo grupo genético reportado previamente (Perez-Enriquez *et al.*, 2018c).

#### **6.1.2.2 Análisis *in silico* para evaluar reacción multiplex para la obtención de genotipos por GT-seq**

Se evaluó incluir los SNPs seleccionados para asignación poblacional y análisis de parentesco, así como un SNP para identificación de sexo (Perez-Enriquez *et al.*, 2020) en una misma plataforma de obtención de genotipos. La genotipificación por GT-seq (Campbell *et al.*, 2015) consiste en la secuenciación de productos de una PCR multiplex, la cual puede incluir hasta centenas de oligos en una misma reacción.

Se diseñaron oligos sentido y antisentido con Primer3 (Untergasser *et al.*, 2012) para todos los SNPs seleccionados, incluyendo el SNP de identificación de sexo (Perez-Enriquez *et al.*, 2020); se

consideraron temperaturas de fusión ( $T_m$ ) homogéneas (58°C min, 60°C óptima, 63°C máx.). En función a los requerimientos de la técnica de GT-seq, a la secuencia de los oligos locus-específicos se les adicionó la secuencia “*small RNA primer*” de Illumina (CGACAGGTTTCAGAGTTCTACAGTCCGACGATC) al oligo sentido y la secuencia “*Read2 primer*” de Illumina (GTGACTGGAGTTTCAGACGTGTGCTCTCCGATCT) al oligo antisentido.

Incluir diversos oligos en una misma reacción multiplex, puede ocasionar interacciones entre ellos que producen artefactos no deseados en la PCR (p.ej. dímeros de oligos, quimeras). Por ello, se utilizó el programa PrimerPooler (Brown *et al.*, 2017) para estimar los valores de delta G (dG) para la potencial hibridación inter e intra-oligos a una temperatura de 60°C y una concentración de cationes monovalentes de 50 mM. Los oligos con interacciones de  $dG \geq 3$  fueron eliminados o rediseñados. Se revisó la alineación de los oligos al genoma de referencia (Zhang *et al.*, 2019) para verificar que se acoplaban a un sitio único y que no existiera traslape entre amplicones; del total de scaffolds (porciones del genoma reconstruido, compuesto de una serie de contigs) del genoma de referencia, se calculó el porcentaje de éstos que contenía al menos un SNP del panel final. Además, se evaluó la posibilidad de utilizar todos los oligos en una misma reacción o dividirlos en 2 o más reacciones distintas.

## **6.2 Caracterización genómica de poblaciones de cultivo de *Penaeus vannamei* en México con un panel de baja densidad de SNPs y validación para análisis de asignación poblacional**

Se realizaron análisis *in silico* para validar un panel de marcadores moleculares con los SNPs obtenidos de la base de datos pública (sección 6.2.1). Para la obtención de genotipos, se evaluó utilizar HRM (curvas de disociación de alta resolución; sección 6.2.2) y GT-seq (sección 6.3.2). Posteriormente, se realizó la caracterización genética de lotes de cultivo en México y se validó el panel de SNPs para análisis de asignación poblacional con muestras ciegas y de origen conocido (sección 6.2.3).

### **6.2.1 Evaluación *in silico* de panel de SNPs para asignación poblacional y estructura genética**

Con el panel de SNPs seleccionados para estructura genética y asignación poblacional (sección 6.1.2.1), se realizaron análisis *in silico* con los datos de librerías secuenciadas de *P. vannamei* (Perez-Enriquez *et al.*, 2018c) para determinar el número mínimo de marcadores que integrarían

el panel final de SNPs. Del archivo vcf del filtrado de SNPs con vcftools (Danecek *et al.*, 2011) (sección 6.1.2.1), se extrajo en Unix (comando grep) los genotipos de los SNPs seleccionados. El archivo se transformó con PGDSpider 2.1.1.5 (Lischer y Excoffier, 2012) en los formatos requeridos para los programas posteriormente utilizados.

Los análisis de asignación de población, que consisten en estimar la probabilidad que un conjunto de individuos pertenezca a una o más grupos de referencia, se realizaron con GeneClass 2 (Piry *et al.*, 2004) con el método bayesiano (Rannala y Mountain, 1997). La población de referencia consistió en 80 individuos pertenecientes a 5 grupos genéticos en México, cuyo origen fue determinado previamente (Perez-Enriquez *et al.*, 2018c). Para cada individuo, se obtuvo un score de pertenecer a cada uno de los 5 grupos genéticos posibles (la suma de todos los scores equivale al 100%). Se consideró como asignación correcta, los individuos con un score mayor o igual del 90% de pertenecer a un grupo. Se calculó el porcentaje de asignación total y se compararon las asignaciones obtenidas con la previamente asignada en la población de referencia para calcular el porcentaje de asignación correcta. Se realizó una serie de análisis variando el número de marcadores de forma descendente con 35, 30, 25, 21, 19, 16 y 15 SNPs. Para verificar la estructura genética, se realizó un Análisis Discriminante de Componentes Principales (DAPC) con Adegnet (Jombart, 2008) utilizando 21 SNPs.

Un análisis adicional de asignación poblacional fue realizado con GeneClass 2 (Piry *et al.*, 2004), en el cual se incluyó el panel de SNPs para GT-seq (tanto los SNPs seleccionados para asignación poblacional, así como los SNPs seleccionados para análisis de parentesco; sección 6.1.2.2). De la población de referencia, se removió el 30% de los individuos al azar y los individuos removidos fueron utilizados como las muestras de “origen desconocido” para ser asignadas. Este proceso se repitió por triplicado. De igual manera, se realizó un Análisis Discriminante de Componentes Principales (DAPC) con Adegnet (Jombart, 2008).

### **6.2.2 Estandarización de obtención de genotipos mediante curvas de disociación de alta resolución (HRM).**

Para cada uno de los SNPs seleccionados para estructura genética y asignación poblacional en la sección 6.1.2.1, se diseñaron oligos sentido y antisentido con el programa Primer3 (Untergasser *et al.*, 2012) para obtener un fragmento de 50-120 pb en el cual se encuentre un solo SNP. Se

calcularon las temperaturas de alineamiento ( $T_a$ ) con la regla de Wallace para obtener la temperatura de fusión ( $T_m$ ), a la cual se restaron  $5^\circ\text{C}$  para obtener la  $T_a$ :

$$2(A + T) + 4(G + C) \quad (1)$$

Para verificar funcionalidad de los cebadores y optimizar la temperatura de alineamiento, se amplificaron 8 muestras de ADN por PCR en un termociclador C1000 touch (BioRad). La mezcla de reacción fue de 11  $\mu\text{l}$  de volumen final que contenía 1  $\mu\text{l}$  de ADN (30  $\text{ng}/\mu\text{l}$ ) y una concentración final de 1.5 mM de  $\text{MgCl}_2$ , 0.2 mM de dNTPs, 0.2  $\mu\text{M}$  de oligos sentido y antisentido, 1x de buffer para Taq polimerasa y 0.03 U/ $\mu\text{l}$  de Taq polimerasa GoTaq (Promega). Las condiciones de la amplificación fueron: desnaturalización inicial de  $95^\circ\text{C}$  por 5 min; 30 ciclos de desnaturalización a  $95^\circ\text{C}$  por 30 s, temperatura de alineamiento por 30 s y extensión a  $72^\circ\text{C}$  por 45 s; extensión final a  $72^\circ\text{C}$  por 10 min. Los productos de PCR se visualizaron en geles de agarosa al 1%. Se seleccionaron los SNPs en los que se obtuvo un fragmento único del tamaño esperado.

La obtención de genotipos de cada SNP se realizó mediante el análisis de curvas de disociación de alta resolución o HRM por sus siglas en inglés (*High Resolution Melt*), el cual se realizó en un termociclador tiempo real CFX96 (BioRad).

Se realizó la extracción de ADN de 24 muestras de músculo y/o pleópodo (12 individuos de cultivo de tres orígenes distintos y 12 individuos silvestres) con el kit de purificación SV 96 Wizard (Promega) siguiendo las instrucciones del fabricante, hasta el paso de la elución del ADN, el cual se hizo con 250  $\mu\text{l}$  de agua libre de nucleasas. El ADN obtenido se cuantificó con un espectrofotómetro Nanodrop 2000 (Thermo Scientific).

Se analizaron inicialmente ocho muestras de ADN (cuatro de individuos de cultivo y cuatro silvestres) por triplicado. La mezcla de reacción fue de 15  $\mu\text{l}$  de volumen final que contenía 5  $\mu\text{l}$  de ADN (8  $\text{ng}/\mu\text{l}$ ) y una concentración final de 1.5 mM de  $\text{MgCl}_2$ , 0.2 mM de dNTPs, 0.2  $\mu\text{M}$  de cebador sentido y antisentido, 1x de EvaGreen (Biotium), 1x de buffer para Taq polimerasa y 0.04 U/ $\mu\text{l}$  de Taq polimerasa GoTaq (Promega). Las condiciones de la qPCR-HRM fueron: desnaturalización inicial a  $95^\circ\text{C}$  por 3 min; 35 ciclos a  $95^\circ\text{C}$  por 10 s, temperatura de alineamiento por 30 s y  $72^\circ\text{C}$  por 30 s; seguido del análisis de la curva de disociación de  $65^\circ\text{C}$ - $95^\circ\text{C}$



incrementando gradualmente 0.1°C con 5 s de lectura de fluorescencia de la placa. Se analizaron 24 muestras por triplicado por qPCR-HRM para optimizar las curvas y obtención de genotipos.

El software del termociclador CFX96 genera los datos de la fluorescencia en RFU (Unidades de Fluorescencia Relativa) en función de la temperatura durante el análisis de la curva de disociación, además de la temperatura inicial y final de cada curva de disociación y la derivada de la fluorescencia en función de la temperatura. Los datos fueron exportados a Excel donde inicialmente fueron graficados considerando el rango de temperatura de la curva dado por el equipo. Al observar la gráfica, se acotó el rango de temperatura del punto previo al descenso de la fluorescencia hasta que desciende completamente. Este rango representa del punto previo al inicio de la disociación de los fragmentos de ADN hasta el punto en que todos se encuentran separados. Este intervalo de datos fue usado en los análisis subsecuentes.

Se generaron tres curvas por cada SNP: (1) una curva normalizada, y a partir de esta se generó (2) una curva diferenciada; con la derivada de la fluorescencia se realizó (3) la curva derivativa. En el anexo C se detalla la descripción y cálculo de las curvas.

### **6.2.3 Caracterización genómica de poblaciones de cultivo y análisis de asignación poblacional**

Se tomaron muestras de pleópodo de camarones (n=203) provenientes de 14 lotes de cultivo (n=168) provenientes de distintos laboratorios de producción del noroeste de México, así como de poblaciones silvestres provenientes de tres sitios de colecta (n=35; Sinaloa, Sinaloa Norte y Guerrero). Las muestras de cultivo fueron procesadas como muestras ciegas, es decir, al momento de analizarlas se desconocía el origen y posteriormente los resultados se cotejaron con los orígenes reportados por los laboratorios de producción.

Se realizó la extracción de ADN de las muestras de tejido con el kit DNeasy blood and tissue (Qiagen) siguiendo las instrucciones del fabricante, se cuantificó con un espectrofotómetro Nanodrop 2000 y se prepararon alícuotas a 25 ng/μl. Se obtuvieron los genotipos adaptando el método de GT-seq (ver sección 6.3.2), SNPs e individuos con genotipos faltantes >50% fueron descartados.

Los lotes de cultivo se analizaron con el programa Structure (Pritchard *et al.*, 2000), incluyendo individuos (n=80) pertenecientes a cinco grupos genéticos de origen conocido previamente reportados (Perez-Enriquez *et al.*, 2018c), estos se usaron como muestras de referencia para inferir el origen de las muestras ciegas. Se corrió el programa con 100,000 pasos de calentamiento y 100,000 repeticiones de Markov para K=1 a K=20 y 5 réplicas para cada K. La K más probable fue calculada por el método de Evanno con Structure Harvester (Earl y VonHoldt, 2012) y CLUMPAK (Kopelman *et al.*, 2015). Se realizó un segundo análisis incluyendo las muestras silvestres, con los mismos parámetros arriba mencionados para K=1 a K=25. Además, se hizo un Análisis Discriminante de Componentes Principales (DAPC) con Adegnet (Jombart, 2008) para verificar estructura poblacional.

Se realizó un análisis de asignación poblacional con GeneClass2 (Piry *et al.*, 2004) con método bayesiano (Rannala y Mountain, 1997), en el cual se utilizaron los datos de Perez-Enriquez *et al.* (2018c) como población de referencia. Asignaciones con un score  $\geq 99\%$  fueron contabilizadas; los resultados de las asignaciones se compararon con el origen reportado (dato proporcionado por el laboratorio de producción) y se calculó el porcentaje de asignaciones correctas.

### **6.3 Validación de panel de SNPs de baja densidad para análisis de parentesco en cultivo de camarón blanco *P. vannamei***

Se realizó una evaluación *in silico* de los SNPs seleccionados para GT-seq de la sección 6.1.2, que incluyó tanto los SNPs de parentesco, como los de asignación poblacional (sección 6.3.1), bajo la premisa de analizar todos los SNPs con un mismo método de obtención de genotipos; se adaptó el método de GT-seq para la obtención de genotipos (sección 6.3.2); se analizaron muestras de camarón de padres y progenie de origen conocido en análisis parentesco para la validación final del panel (sección 6.3.3).

#### **6.3.1 Evaluación *in silico* de panel de SNPs para análisis de parentesco**

Para evaluar el panel de SNPs seleccionado para GT-seq (sección 6.1.2), se realizó un análisis *in silico* utilizando los datos de las secuencias de 95 camarones previamente publicadas (Perez-Enriquez *et al.*, 2018c), los cuales se utilizaron también para el desarrollo del panel. Del archivo vcf resultante del filtrado de SNPs (sección 6.1.2.1) con vcftools (Danecek *et al.*, 2011), se

extrajeron en Unix los genotipos de los SNPs seleccionados. El archivo vcf con los datos de dichos marcadores, se transformó con PGDSpider 2.1.1.5 (Lischer y Excoffier, 2012) para poder ser editado según las especificaciones de los programas posteriores.

Se realizó la simulación de 100,000 individuos (progenie) en base a frecuencias alélicas de los SNPs seleccionados utilizando Cervus 3.0.3 (Kalinowski *et al.*, 2007), con 200 padres de sexo desconocido (100 hembras y 100 machos) y el supuesto del 100% y el 90% de padres muestreados.

### **6.3.2 Estandarización de GT-seq para obtención de genotipos**

Muestras de pleópodo de camarón blanco *P. vannamei* (n=333) se utilizaron para estandarizar la obtención de genotipos por GT-seq (Campbell *et al.*, 2015) realizando modificaciones al protocolo. Se realizó la extracción de ADN con el kit DNeasy blood & tissue (Qiagen) siguiendo las instrucciones del fabricante. El ADN fue cuantificado en un Nanodrop 2000 y se verificó que su pureza fuera adecuada (proporción 260/280 de 1.8-2.2 y proporción 260/230  $\geq 1.8$ ). Se prepararon alícuotas a 25 ng/ $\mu$ l.

GT-seq (*Genotyping-in-Thousands by sequencing*; genotipificación en miles por secuenciación) es un método de obtención de genotipos que consiste en la secuenciación de productos de una PCR multiplex, en la reacción se amplifican los fragmentos que contienen cada uno de los SNPs.

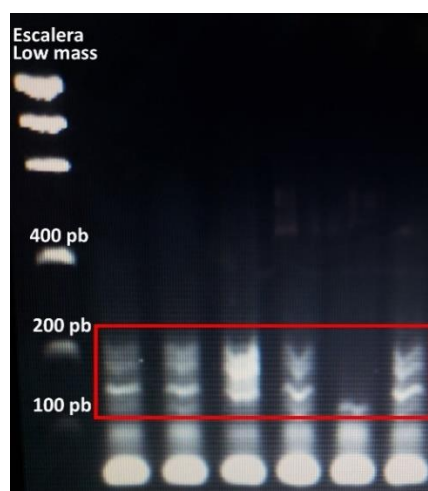
Como se mencionó en la sección 6.1.2.2, se evaluó *in silico* la reacción multiplex que contenía los oligos diseñados para los SNPs seleccionados para asignación poblacional, análisis de parentesco y un SNP de identificación de sexo (Perez-Enriquez *et al.*, 2020), con el objetivo de evitar interacciones no deseadas inter e intra-oligos. Posteriormente, los oligos seleccionados fueron sintetizados en pool con el servicio *Opools* de Integrated DNA Technologies (IDT).

Cabe mencionar que inicialmente se prepararon las librerías de GT-seq siguiendo principalmente el protocolo publicado (Campbell *et al.*, 2015) y posteriormente, se realizaron modificaciones para optimizar los resultados de la secuenciación. La preparación de las librerías consistió en: 1)

PCR1 para amplificación con oligos locus-específicos, 2) PCR2 para la adición de códigos de barra a los fragmentos de la PCR1, 3) Normalización, purificación y cuantificación de librería final, y 4) secuenciación y análisis de datos.

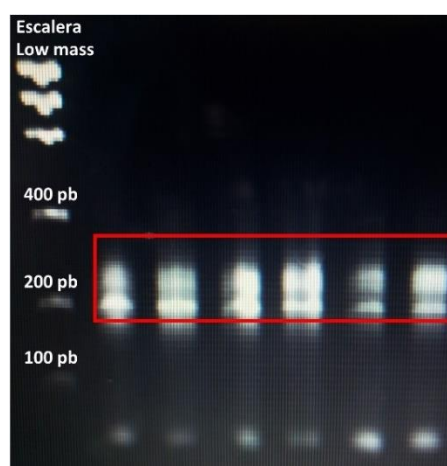
A continuación, el protocolo de preparación de librerías modificado:

1) *Primera PCR (PCR1)*. La amplificación por PCR consistió en dos etapas. La primera (PCR1), es una PCR multiplex que incluye el pool de oligos locus-específicos (Anexo E) que contiene la extensión de secuencias de illumina, secuencia “*small RNA primer*” (CGACAGGTTTCAGAGTTCTACAGTCCGACGATC) al oligo sentido y la secuencia “*Read2 primer*” (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT) al oligo antisentido. La reacción de 7  $\mu$ l de volumen final contiene: 2  $\mu$ l ADN (25 ng/ $\mu$ l), 1  $\mu$ l de pool de oligos (0.125 pmol/ $\mu$ l/oligo), 3.5  $\mu$ l de master mix PCR multiplex (kit Multiplex PCR plus de Qiagen), y 0.5  $\mu$ l de agua MilliQ. Para la amplificación se utilizó un protocolo *touchdown* con las siguientes condiciones: 95°C por 10 min; 8 ciclos de 95°C por 30 s, 62-58°C por 1 min (-0.5°C por ciclo), 72°C por 1 min; 15 ciclos de 95°C por 30 s, 65°C por 30 s, 72°C por 1 min; y 72°C por 2 min como extensión final. Para verificar que la amplificación fue exitosa y que se obtuvieran los fragmentos esperados (126-216 pb), muestras tomadas al azar de la PCR1 se visualizaron en gels de agarosa al 2%; en la Fig. 5 se observa un ejemplo de un gel de agarosa en el cual se señalan los fragmentos esperados en un recuadro rojo. Los productos de la PCR1 fueron purificados posteriormente con 1.3 $\times$  de perlas magnéticas AMPure XP beads (Beckman Coulter).



**Figura 5.** Visualización de fragmentos de PCR1 en gel de agarosa al 2%. Fragmentos esperados de 126-216 pb señalados en el rectángulo rojo.

2) *Segunda PCR (PCR2)*. La segunda etapa (PCR2) consiste en el indexado de los fragmentos de la PCR1 con doble código de barras, oligos i5 e i7 (Anexo F). La reacción de la PCR2 fue de 10  $\mu$ l de volumen final: 3  $\mu$ l de PCR1, 1  $\mu$ l de oligo i5 (10 pmol/ $\mu$ l), 1  $\mu$ l de oligo i7 (10 pmol/ $\mu$ l), y 5  $\mu$ l de master mix PCR multiplex (kit Multiplex PCR plus de Qiagen). Las condiciones de amplificación fueron las siguientes: 95°C por 15 min; 10 ciclos de 98°C por 10 s, 65°C por 30 s, 72°C por 30 s; 72°C por 5 min de extensión final. Para verificar que la amplificación fue exitosa y que se obtuvieran los fragmentos esperados (191-281 pb), muestras tomadas al azar se visualizaron en geles de agarosa al 2%; en la Fig. 6 se muestra un ejemplo de un gel de agarosa, en el cual se señalan los fragmentos esperados con un recuadro rojo.



**Figura 6.** Visualización de fragmentos de PCR2 previo a la normalización y purificación. Fragmentos esperados de 191-281 pb resaltados en el rectángulo rojo.

3) *Normalización, purificación y cuantificación*. El producto de la PCR2 se normalizó con el kit de normalización en placa SequalPrep (Thermo Fisher) siguiendo las instrucciones del fabricante. Se preparó una mezcla por placa de las librerías individuales, tomando 10  $\mu$ l por cada pozo de la placa y se mezclaron en un tubo de 1.5ml. Las librerías por placa se purificaron con 1.8 $\times$  de perlas magnéticas AMPure XP beads (Beckman Coulter) y se cuantificaron por duplicado con Qubit. Las librerías por placa fueron normalizadas a 2 nM y mezcladas en una librería final.

4) *Secuenciación y análisis de datos*. Previo a la secuenciación, la librería final fue desnaturalizada y diluida a 10 pM. La secuenciación se realizó en un Miseq Illumina en el Laboratorio de Genómica y Bioinformática del CIBNOR, con una celda nano v2 de 300 ciclos y 1 millón de lecturas. Se descargaron los archivos Fastq generados de la secuenciación y se analizó la calidad de las lecturas

con FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Se eliminaron las secuencias de baja calidad y los adaptadores con el programa Trimmomatic (Bolger *et al.*, 2014). Las secuencias recortadas fueron analizadas con Perl utilizando el script desarrollado por Campbell *et al.* (2015) para la obtención de genotipos, el cual fue modificado para conservar los genotipos con profundidad de lectura  $\geq 4\times$ , en el anexo G se incluye dicho script, así como el archivo de entrada necesario para la obtención de genotipos en el anexo H.

### 6.3.3 Validación de panel de SNPs para análisis de parentesco

Para la validación en análisis de parentesco del panel de SNPs seleccionado, se utilizaron muestras de tejido de camarón pertenecientes a 21 familias de hermanos completos ( $n=130$ ; 21 hembras, 21 machos y 88 progenie), estas familias fueron previamente identificadas con un panel comercial (Nolasco-Alzaga *et al.*, 2018). Se realizó la extracción de ADN con el kit DNeasy blood and tissue (Qiagen) siguiendo las instrucciones del fabricante, se cuantificó con un Nanodrop 2000 y se prepararon alícuotas a 25 ng/ $\mu$ l. Se obtuvieron los genotipos con el método de GT-seq (ver sección anterior), SNPs e individuos con genotipos faltantes  $>50\%$  fueron descartados.

Se realizaron análisis de paternidad para ambos padres con sexo conocido en Cervus 3.0.7 (Kalinowski *et al.*, 2007). Los análisis de parentesco con este programa consisten en tres pasos: cálculo de frecuencias alélicas, simulación y asignación de la progenie. En la simulación se utilizaron los siguientes parámetros: simulación de 100,000 individuos de la progenie, 0.90 y 1.0 de proporción de padres muestreados, 0.10 de proporción de genotipos faltantes y de error en los cálculos de *likelihood*. La asignación de la progenie a ambos padres se realizó con el 90% y 95% de nivel de confianza.

Además, se realizaron análisis de parentesco con Colony 2.0.6.8 (Jones y Wang, 2010) para asignación de la progenie a ambos padres, así como asignación familiar. Los parámetros utilizados en el análisis fueron: 0.01 de error en los genotipos para cada marcador, 90% de probabilidad de presencia de los padres en los candidatos, método de *full-likelihood* con los parámetros de default y se realizaron 5 corridas del análisis.

Los resultados obtenidos se compararon con las asignaciones previamente realizadas (Nolasco-Alzaga *et al.*, 2018) con un panel comercial y se determinaron los porcentajes de asignación correcta.

## 7. RESULTADOS

### 7.1 Desarrollo de paneles de marcadores moleculares tipo SNP en camarón blanco *P. vannamei*

#### 7.1.1 Panel de SNPs con 2bRAD

El número de sitios de reconocimiento para las cuatro enzimas IIB, identificados *in silico* en el genoma de camarón, se presenta en la Tabla 2. Para este trabajo, parte del criterio de selección de la enzima fue que el número de fragmentos fuera reducido para evaluar la posibilidad de obtener un panel de SNPs de baja densidad (decenas de SNPs), por lo que la enzima BsaXI se descartó por tener el mayor número de sitios de reconocimiento (344,380). Aunque la enzima CspCI tuvo el menor número de sitios identificados (60,807), la variación en el número de bases de corte (5' <sub>10-11</sub>(N)CAA(N)<sub>5</sub>GTGG(N)<sub>12-13</sub> 3') provoca una incertidumbre no deseada durante la digestión (New England Biolabs, comunicación personal). Por lo que, se eligió la enzima Bcgl.

**Tabla 2.** Número de sitios de reconocimiento de enzimas IIB en genoma de camarón

Enzima	Número de sitios de reconocimiento		
	Sentido	Antisentido	Total
Bcgl	106,530	107,988	214,518
BsaXI	205,589	138,791	344,380
Alfi <sup>1</sup>	136,798	NA	136,798
CspCI	31,678	29,129	60,807

<sup>1</sup>La secuencia de reconocimiento en sentido y antisentido es la misma.

El número de sitios remanentes si incluyéramos bases selectivas en los adaptadores (Tabla 3) fue bastante homogéneo para las cuatro enzimas, por lo que no fue en criterio que influyera para escoger la enzima Bcgl en la preparación de librerías 2bRAD. Pero si, para considerar utilizar por lo menos dos bases selectivas en los adaptadores.



**Tabla 3.** Porcentaje de fragmentos remanentes al incluir adaptadores con bases selectivas

Enzima	Sitios de reconocimiento	Porcentaje de sitios remanentes				
		1 base	2 bases	3 bases	4 bases	5 bases
Bcgl	214,518	23.0%	5.5%	1.3%	0.3%	NA
BsaXI	344,380	22.9%	5.5%	1.3%	0.3%	0.1%
AlfI	136,798	23.0%	5.5%	1.3%	0.3%	NA
CspCI	60,807	22.3%	5.2%	1.3%	0.3%	NA

En el análisis *in silico* para estimar que proporción de fragmentos tendrían SNPs si utilizamos la enzima Bcgl, se obtuvieron 1,534 SNPs en 234 scaffolds del genoma; en 50 fragmentos revisados, al menos el 65% tuvo SNP dentro del fragmento esperado con la enzima Bcgl. Por lo que, si consideramos este resultado, podríamos esperar que de 214,518 sitios de reconocimiento (Tabla 2) se podrían obtener 139,437 SNPs aprox. Si incluimos una, dos, tres y cuatro bases selectivas (Tabla 3) se podrían esperar aprox. 32,071, 7,669, 1,813 y 418 SNPs, respectivamente. Tomando en consideración lo anterior, se prepararon las librerías 2bRAD con 2 y 4 bases selectivas.

Los resultados de la secuenciación de las cuatro librerías se presentan en la Tabla 4. El número de lecturas fue homogéneo excepto por la librería del pool silvestre con dos bases selectivas, que presentó mayor número de lecturas. El porcentaje de alineamiento de las secuencias al genoma de referencia fue similar en las librerías del pool de cultivo y silvestre, pero ligeramente mayor en las muestras silvestre en comparación con las muestras de cultivo.

El número de SNPs obtenido utilizando dos bases selectivas fue de 19,105 SNPs y utilizando cuatro bases selectivas fue de 2,874 SNPs después del filtrado.

**Tabla 4.** Número de lecturas y porcentaje de alineamiento de librerías 2bRAD

<b>Librería</b>	<b>Número de lecturas</b>	<b>% alineamiento</b>
Pool cultivo con 2 bases selectivas	1,110,928	64.4
Pool silvestre con 2 bases selectivas	4,532,602	71.9
Pool cultivo con 4 bases selectivas	1,893,650	63.4
Pool silvestre con 4 bases selectivas	1,341,554	72.0

### 7.1.2 Panel de SNPs utilizando datos públicos de secuenciación

#### 7.1.2.1 Análisis bioinformático para obtención de SNPs

Se obtuvieron 14,268,805 variantes, previo a los procesos de filtrado. Después del filtrado de SNPs para asignación poblacional y estructura genética, se obtuvieron 16,991 SNPs; posteriormente, se seleccionaron 2,050 SNPs por tener una región conservada de 30 bases a cada lado, los cuales fueron utilizados en los análisis de *Fst* pareados. Finalmente, se seleccionaron 46 SNPs con los valores de diferenciación más altos en los análisis pareados ( $Fst > 0.20$ ).

En el caso de los SNPs para análisis de parentesco, después del filtrado, se obtuvieron 25,286 SNPs; posteriormente, se seleccionaron 9,784 SNPs por tener presuntamente 30 bases de región conservada a cada lado. Finalmente se seleccionaron 132 SNPs por tener máximo otra variante más en el fragmento secuenciado de 150 bases.

La prueba de desequilibrio de ligamiento realizada con el panel final de 118 SNPs (ver siguiente sección), no mostró valores significativos después de la corrección secuencial de Bonferroni ( $p=0.0000074$ ), por lo que los SNPs no se encuentran ligados entre sí.

### 7.1.2.2 Análisis *in silico* para evaluar reacción multiplex

Se logró diseñar oligos locus-específicos para: 19 SNPs seleccionados para estructura genética y asignación poblacional, 116 SNPs seleccionados para análisis de parentesco, y un SNP para identificación de sexo.

De las posibles interacciones entre e intra-oligos, 23 tuvieron un  $dG \geq 3$ ; debido a esto 12 SNPs (24 pares de oligos) fueron eliminados, ya que interactuaban con varios oligos y/o no era posible rediseñarlos. Los resultados del alineamiento al genoma de referencia mostraron que 12 pares de oligos produjeron más de un amplicón, al hibridarse en más de un sitio del genoma, por lo que se descartaron 6 SNPs más del panel.

El panel final quedó conformado por 118 SNPs (100 SNPs para análisis de parentesco, 17 para estructura genética y asignación poblacional, y un SNP para identificación de sexo; anexo E).

El alineamiento de los oligos de estos 118 SNPs al genoma de referencia, mostró que no existía traslape entre los amplicones esperados. Además, de que se encontraban en 88 scaffolds (al menos un SNP) de los 4683 reportados en el genoma de referencia, lo que equivale al 1.88%. Cabe mencionar que los oligos de 25 SNPs no presentaron hibridación al genoma de referencia, lo cual puede deberse a que el genoma de referencia utilizado está incompleto.

Finalmente, los 236 oligos de los 118 SNPs pueden ser utilizados en una misma reacción multiplex, sin la necesidad de dividir la reacción según los resultados del análisis con PrimerPooler, esto ya considerando las extensiones que se deben de adicionar a los oligos para su obtención de genotipos mediante GT-seq.

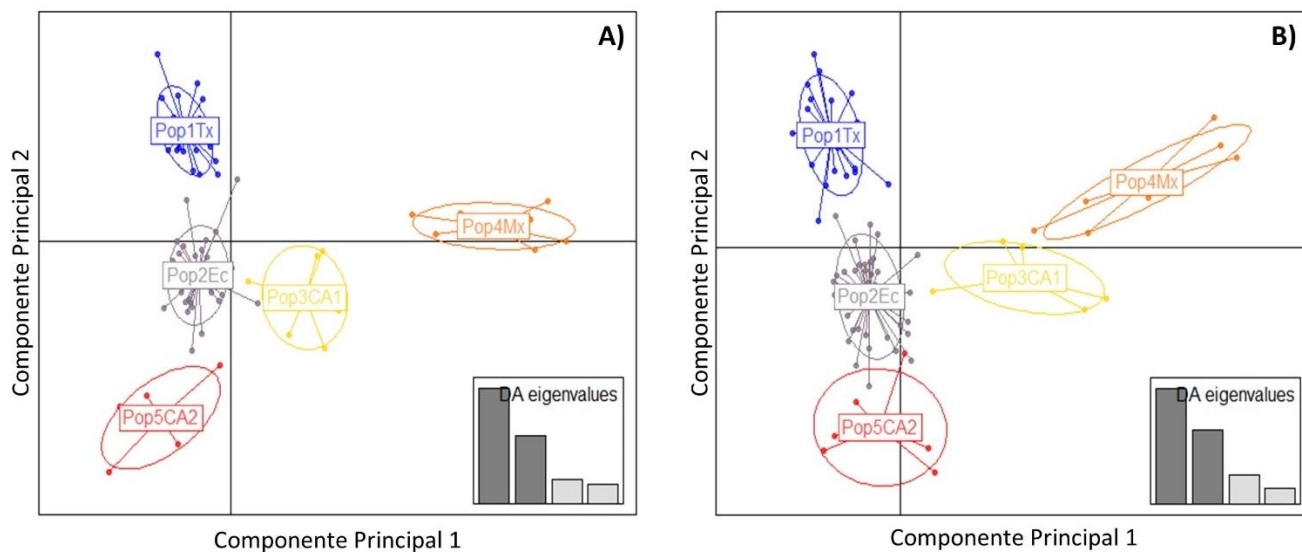
## 7.2 Caracterización genómica de poblaciones de cultivo de *P. vannamei* en México con un panel de baja densidad de SNPs y validación para análisis de asignación poblacional

### 7.2.1 Análisis *in silico* para validar panel de SNPs para asignación poblacional y estructura genética

En los resultados de asignación poblacional con diferente número de SNPs (Tabla 5), se encontró que el porcentaje de asignación se mantuvo similar con 35, 30 y 25 marcadores, sin embargo, con 25 SNPs se observa 1.3% de error. Cuando se revisaron los valores de diferenciación poblacional (*Fst*) de los SNPs que conformaban cada panel (datos no mostrados), se observó que la eliminación de SNPs con valores de *Fst* altos para distinguir los grupos genéticos 1 vs 2, 2 vs 3 y 2 vs 5 tenían un efecto mayor en la asignación. Por lo que, en los análisis con 21, 19 y 16 SNPs, se dio prioridad a seleccionar aquellos SNPs con valores de *Fst*  $\geq 0.20$  para distinguir los grupos arriba mencionados. Por ejemplo, para el panel de 21 SNPs se seleccionaron todos los marcadores disponibles del panel inicial (35 SNPs) con valores de *Fst*  $\geq 0.20$  para estos grupos. El porcentaje de asignación con 21 SNPs no presentó errores y sólo disminuyó 3.7% del de 35 SNPs. Considerando estos resultados, y con el objetivo de optimizar el número de marcadores, se dio prioridad a incluir estos 21 SNPs en el panel final para GT-seq (sección 6.1.2.2). Además, en el Análisis Discriminante de Componentes Principales (DAPC), se observó la identificación de los cinco grupos genéticos esperados con el panel de 35 (Fig. 7A) y 21 SNPs (Fig. 7B).

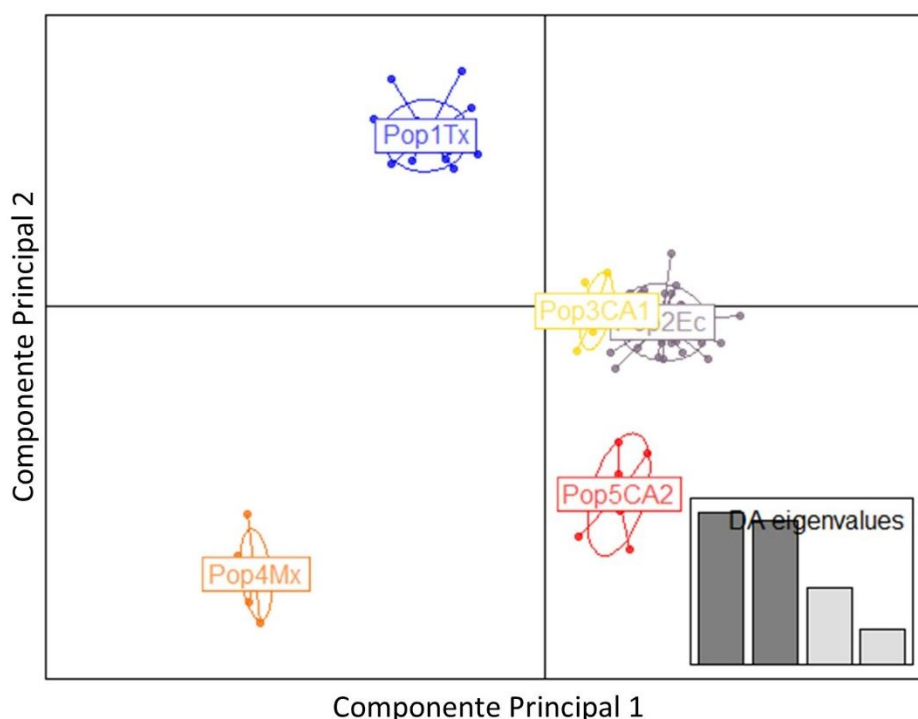
**Tabla 5.** Asignación poblacional con diferente número de SNPs en análisis *in silico*

	35 SNPs	30 SNPs	25 SNPs	21 SNPs	19 SNPs	16 SNPs	15 SNPs
% Asignación	90%	87.5%	87.5%	86.3%	81.3%	63.8%	60%
% Asig. incorrectas	0%	0%	1.3%	0%	1.3%	0%	0%



**Figura 7.** Análisis Discriminante de Componentes Principales (DAPC). A) DAPC con 35 SNPs. B) DAPC con 21 SNPs. Pop1Tx (origen Texas), Pop2Ec (origen Ecuador), Pop3CA1 (origen Centroamérica 1), Pop4Mx (origen México), Pop5CA2 (origen Centroamérica 1). Grupos genéticos reportados por Perez-Enriquez *et al.* (2018c).

En la validación para asignación poblacional del panel final de 117 SNPs para GT-seq (el SNP de identificación de sexo fue excluido del análisis), del 92-96% de las muestras fueron asignadas a su población de origen con un error de asignación del 5-13%. En el DAPC (Fig. 8) se observó una clara separación en cuatro grupos genéticos, sin embargo, el grupo genético 2 y 3 (origen Ecuador y Centroamérica 1, respectivamente) mostraron una ligera mezcla entre ellos.



**Figura 8.** Análisis Discriminante de Componentes Principales (DAPC) con 117 SNPs. Pop1Tx (origen Texas), Pop2Ec (origen Ecuador), Pop3CA1 (origen Centroamérica 1), Pop4Mx (origen México), Pop5CA2 (origen Centroamérica 1). Grupos genéticos reportados por Perez-Enriquez *et al.* (2018c).

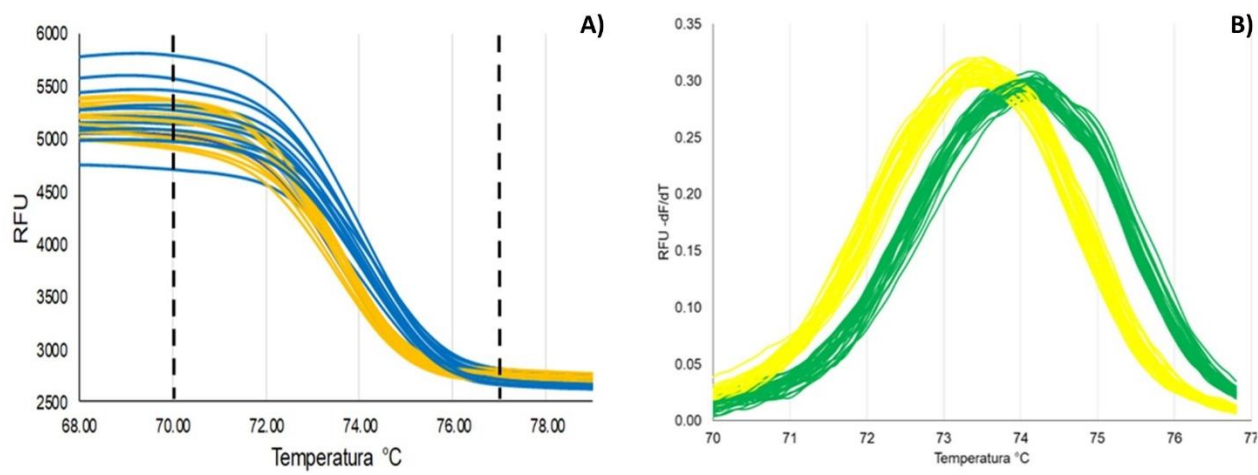
### 7.2.2 Estandarización de obtención de genotipos mediante HRM

Se seleccionaron inicialmente 46 SNPs por sus valores de *Fst* pareados (sección 7.1.2.1), de los cuales fue posible diseñar oligos en 39 SNPs. Una vez probados los oligos por PCR se obtuvo un producto único del tamaño esperado en 37 SNPs.

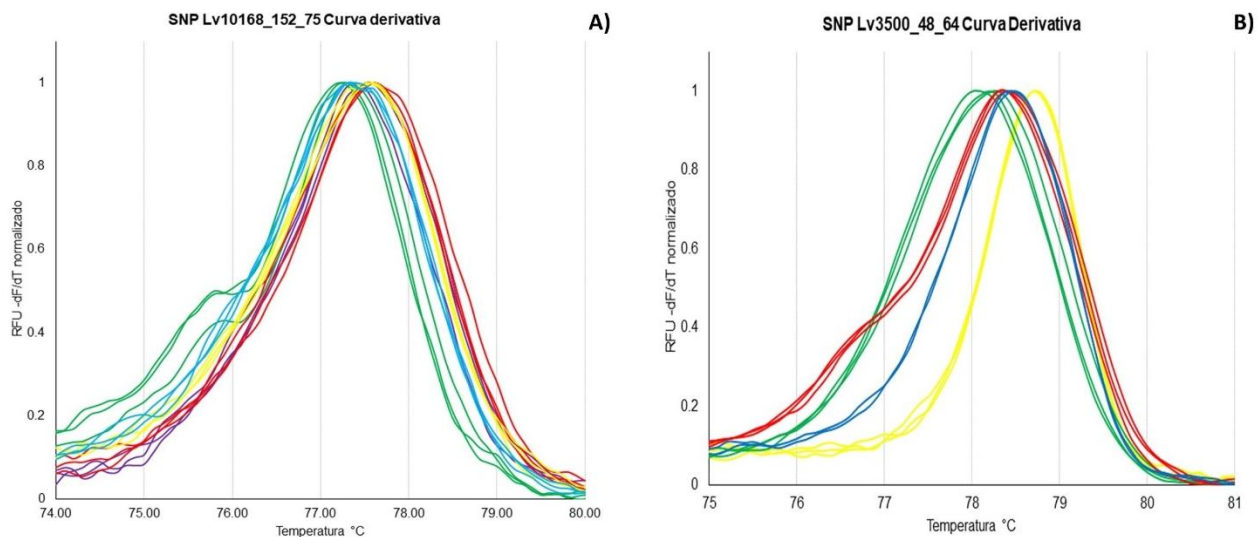
Las temperaturas de alineamiento para cada uno de los SNPs, así como la secuencia de los oligos sentido y antisentido se encuentran en el anexo D.

Se muestra un ejemplo de la curva de disociación y la curva derivativa para uno de los SNPs (Lv24611\_43\_128), en el que se observan claramente homocigotos para ambos alelos (Fig. 9). Sin embargo, en varios casos se obtuvieron curvas con productos inespecíficos, o la diferencia entre un genotipo y otro era poco clara (Fig. 10A), o se detectó la presencia de fragmentos con más de un SNP, lo que resultó en la presencia de 4 curvas distintas (Fig. 10B). Por lo que, los oligos fueron

rediseñados para obtener amplicones menores a 80 pb. En dos SNPs no fue posible rediseñar los oligos por lo que el conjunto final consistió en 35 SNPs.



**Figura 9.** Curvas HRM de SNP Lv24611\_43\_128. A) Curvas de disociación; se señala con las líneas segmentadas los puntos en los que se delimita la curva para su posterior análisis. B) Curva derivativa en el que se observan homocigotos para ambos alelos.



**Figura 10.** Ejemplo de curvas derivativas. Las líneas del mismo color corresponden a un mismo individuo analizado por triplicado. A) SNP donde no se distingue claramente la diferencia entre genotipos. B) SNP donde se observan 4 genotipos distintos.

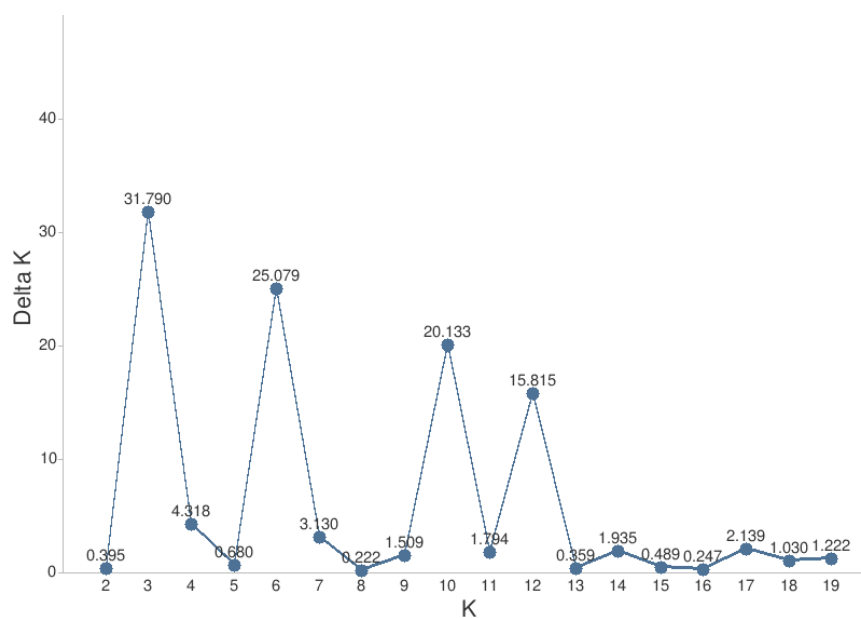
Aunque utilizar HRM es una opción viable para la obtención de genotipos en SNPs, está es más recomendable cuando el número de marcadores es bajo (decenas de SNPs) ya que, aunque es una opción de bajo costo, el tiempo de procesamiento de las muestras es mayor que con otras

opciones que se evaluaron en este trabajo, como GT-seq; por lo que se decidió optar por esta opción para la obtención de genotipos.

### 7.2.3 Caracterización genómica de poblaciones de cultivo y análisis de asignación poblacional

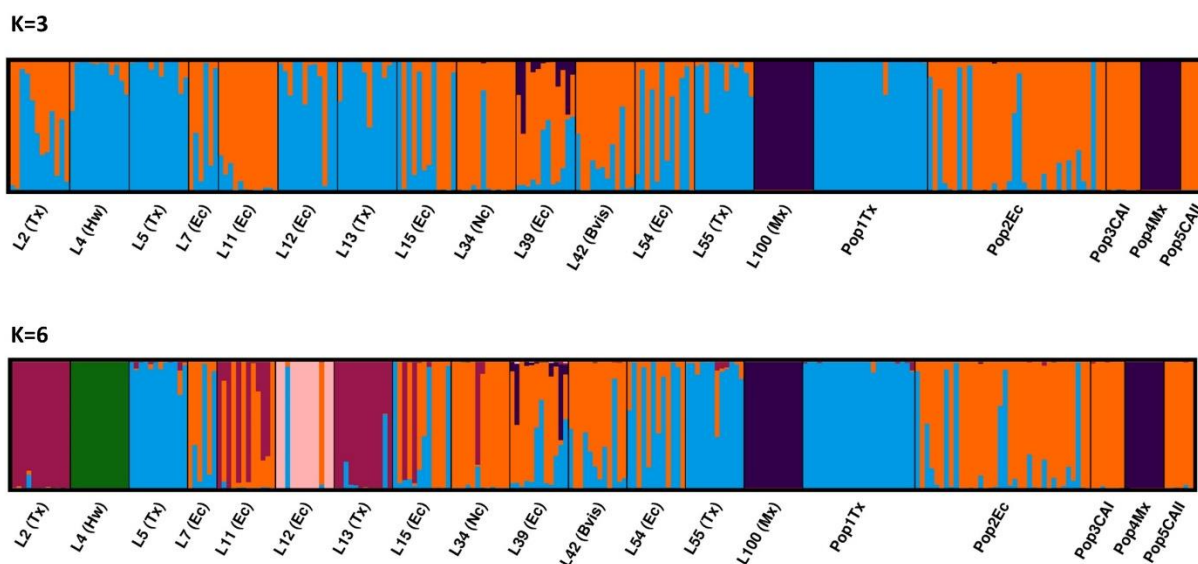
Los resultados de esta sección incluyen el análisis de 197 individuos, ya que 6 muestras fueron descartadas por bajos porcentajes de genotipificación (<50%), así como con 95 SNPs (20 SNPs se descartaron por bajo porcentaje de genotipos).

En el análisis de Structure, que incluye los lotes de cultivo y las muestras de referencia (Perez-Enriquez *et al.*, 2018c), se obtuvo como número de grupos genéticos más probable  $K=3$  seguido de  $K=6$  (Fig.11). En la gráfica de banderas para  $K=6$  (Fig.12), se observa que de los 14 lotes de cultivo analizados como muestra “ciega”, ocho de ellos están claramente definidos con todas o casi todas las muestras de ese lote con proporción de 1.0 de pertenecer a un grupo genético (Tabla 6), mientras que en los demás lotes se aprecia un grado de mezcla, por ejemplo, el Lote 15 y 42.



**Figura 11.** Gráfica de DeltaK para visualizar el número de K más probable ( $K$ =grupos genéticos) incluyendo los 14 lotes de cultivo (muestras ciegas) y las muestras de referencia (Perez-Enriquez *et al.*, 2018c).





**Figura 12.** Gráfica de banderas de Structure para K=3 y K=6. Muestras “ciegas” de 14 lotes de cultivo (L), entre paréntesis origen reportado por el laboratorio de producción; muestras de origen conocido (Pop) reportados por Pérez-Enriquez *et al.* (2018c). Bvis=Buenavista, CAI=Centroamérica1, CAII=Centroamérica2, Ec=Ecuador, Hw=Hawái, Mx=México, Nc=Nicaragua, Tx=Texas.

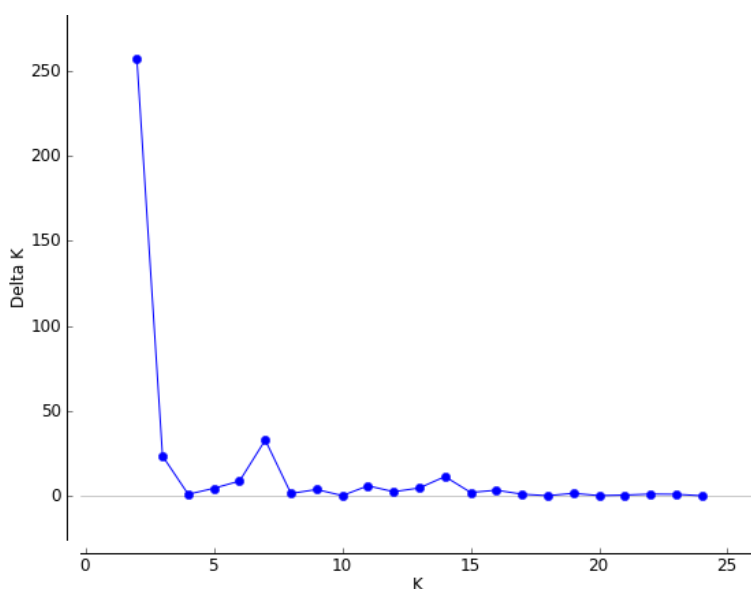
Por otra parte, con los resultados para 6 grupos genéticos (K=6; Fig. 12), se puede inferir que el Lote 100 tiene origen México (morado), y los lotes 5 y 55 tienen como origen Texas (azul), ya que pertenecen al mismo grupo genético de las muestras de origen conocido de Pérez-Enriquez *et al.* (2018c), correspondientes al grupo 2 y 3, con una proporción promedio de membresía >0.90 (Tabla 6); estos resultados coinciden con los orígenes reportados por los laboratorios de producción. De igual manera, considerando los 6 grupos genéticos, los lotes 2 y 13 pertenecen a un mismo grupo (grupo 6), los cuales fueron reportados con origen en Texas, pero son genéticamente distintos al de la muestra de referencia (azul); así como el Lote 4 (verde) el cual pertenece a un grupo genético distinto (grupo 5; Tabla 6) con origen reportado en Hawái, esto último coincide con la gráfica de DAPC (Fig.15B), en la cual se puede observar la separación de este lote de las demás muestras. En esta misma gráfica, se confirma que el lote 100 tiene como origen México ya que se encuentra junto con la muestra de referencia para este origen, así como una clara diferenciación de los lotes que tiene un origen de importación con el de origen en México (Fig. 15).

**Tabla 6.** Origen reportado de lotes de cultivo por laboratorios de producción y origen asignado con análisis de Structure (asignación de grupos genéticos) y GeneClass (asignación de individuos).

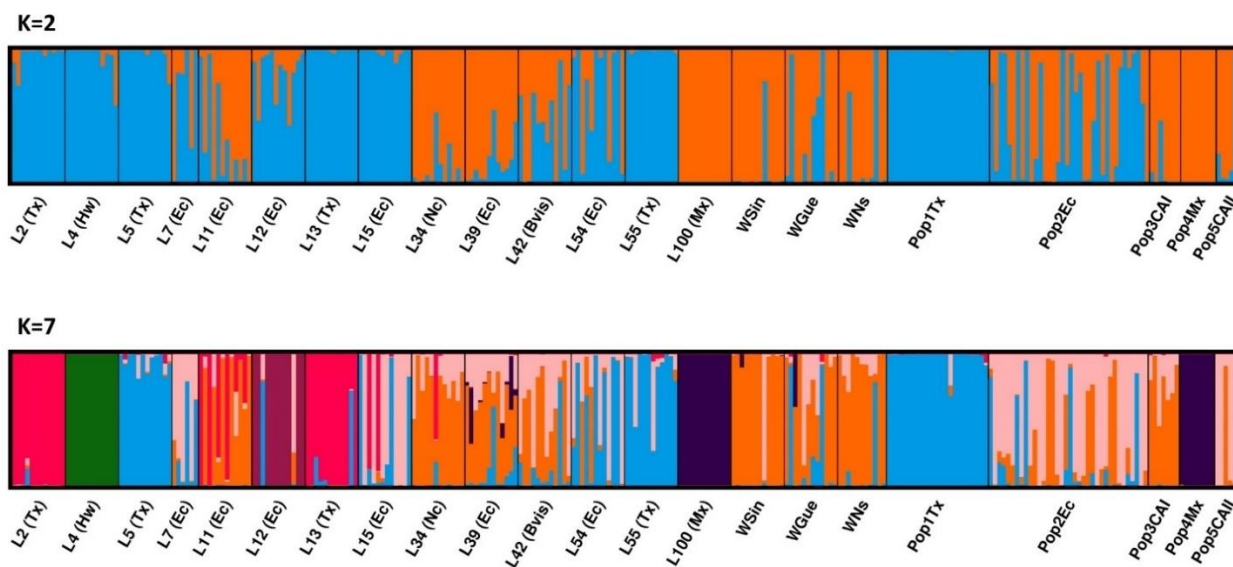
ID Lote	Origen reportado	Grupos asignados Structure <sup>1</sup> (origen inferido)	Origen asignado GeneClass <sup>3</sup> (n° de individuos asignados correctamente)
Lote 2	Texas	Grupo 6	-
Lote 4	Hawái	Grupo 5	-
Lote 5	Texas	Grupo 3 (Texas)	Texas (2)
Lote 7	Ecuador	Mezcla <sup>2</sup>	Ecuador (3)
Lote 11	Ecuador	Mezcla <sup>2</sup>	Ecuador (6)
Lote 12	Ecuador	Grupo 4	Ecuador (2)
Lote 13	Texas	Grupo 6	Texas (5)
Lote 15	Ecuador	Mezcla <sup>2</sup>	Ecuador (6)
Lote 34	Nicaragua	Grupo 1	Centroamérica I (1)
Lote 39	Ecuador	Mezcla <sup>2</sup>	-
Lote 42	Buenavista	Mezcla <sup>2</sup>	-
Lote 54	Ecuador	Mezcla <sup>2</sup>	Ecuador (7)
Lote 55	Texas	Grupo 3 (Texas)	Texas (3)
Lote 100	México	Grupo 2 (México)	México (12)

<sup>1</sup>Proporción promedio de membresía a un grupo >0.80; <sup>2</sup>Proporción de membresía a más de un grupo; <sup>3</sup>Individuos asignados con score >99%.

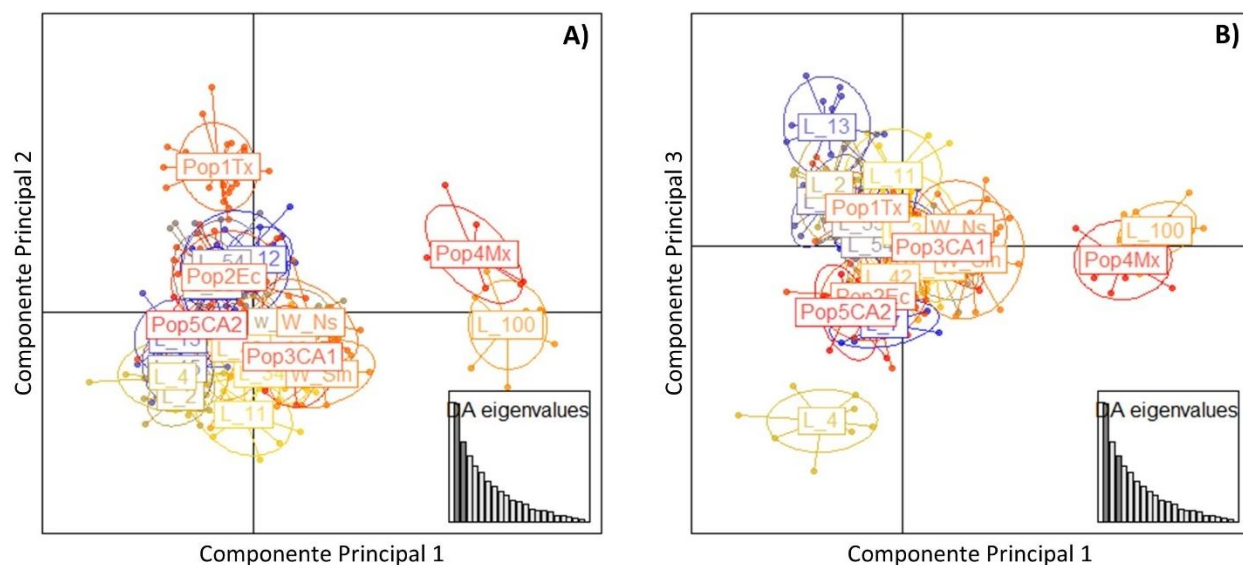
En cuanto a las muestras silvestres, se destaca que no se logran diferenciar genéticamente de algunos lotes de cultivo, como se observa en las gráficas de bandera de Structure con perfiles genéticos similares a los lotes 11, 34, 39 y 42 (Fig. 14) y DAPC (Fig. 15).



**Figura 13.** Gráfica de DeltaK para visualizar el número de K más probable (K=grupos genéticos) incluyendo los 14 lotes de cultivo (muestras ciegas), 3 poblaciones silvestres (Sinaloa, Sinaloa Norte y Guerrero) y las muestras de referencia (Perez-Enriquez *et al.*, 2018c).



**Figura 14.** Gráfica de banderas de Structure para K=2 y K=7. Muestras “ciegas” de 14 lotes de cultivo (L), entre paréntesis origen reportado por el laboratorio de producción; muestras silvestres (W); muestras de origen conocido (Pop) reportados por Pérez-Enriquez *et al.* (2018c). Bvis=Buenavista, CAI=Centroamérica1, CAII=Centroamérica2, Ec=Ecuador, Hw=Hawái, Mx=México, Nc=Nicaragua, Tx=Texas, WSin=Sinaloa, WGue=Guerrero, WNs=Sinaloa noroeste.



**Figura 15.** Gráfica de DAPC (Análisis Discriminante de Componentes Principales) para las muestras “ciegas” de 14 lotes de cultivo (L); muestras silvestres (W), Sin=Sinaloa, Gue=Guerrero, Ns=Sinaloa noroeste; muestras de origen conocido (Pop) reportados por Pérez-Enriquez *et al.* (2018c), Pop1Tx=Texas, Pop2Ec=Ecuador, Pop3CA1=Centroamérica1, Pop4Mx=México, Pop5CA2=Centroamérica2. A) DAPC de Componentes Principales 1 y 2. B) DAPC de Componentes Principales 1 y 3.

Con el análisis de asignación poblacional con GeneClass2, se obtuvo un 46% (n=74) de asignación de las muestras de los lotes de cultivo a las poblaciones de referencia reportadas por Pérez-Enriquez *et al.* (2018c) con 18 asignaciones incorrectas, y 9 individuos del lote 42 reportado como Buenavista (lote de origen silvestre con pocas generaciones en cultivo; Rafael Ruíz, comunicación personal) cuyas muestras fueron principalmente asignadas a la población de referencia de Ecuador; esto coincide con los resultados de Structure (Fig. 12) en el que se observan perfiles genéticos muy similares entre el lote 42 y la muestra de referencia de Ecuador. Las asignaciones poblacionales con GeneClass2 permitieron inferir muestras con origen de los lotes 7, 11 y 15 como Ecuador (que coincide con el reportado por los laboratorios de producción; Tabla 6).

#### 7.2.4 Evaluación *in silico* de panel de SNPs para análisis de parentesco

Para la evaluación *in silico* de 117 SNPs (panel para GT-seq, excluyendo el SNP de identificación de sexo), los resultados de asignación de la progenie simulada a sus padres se presenta en la Tabla 7; considerando la proporción de padres muestreados utilizada en las simulaciones, los resultados de asignación fueron como se esperaban (ejemplo, con el 1.0 de padres muestreados, el 100% de

la progenie asignada), aunque con el 95% de nivel de confianza fue ligeramente bajo (77% de progenie asignada) con el 0.90 de padres de muestreados.

Además, se obtuvo una probabilidad combinada de no-exclusión de  $5.87 \times 10^{-15}$  para la asignación a ambos padres, es decir, la probabilidad promedio de que con este panel de loci, se realice la asignación de una progenie a una pareja de progenitores que no sean sus padres reales.

**Tabla 7.** Porcentaje de asignación en análisis de parentesco *in silico* con 117 SNPs

Proporción de padres muestreados	Nivel de confianza	
	90%	95%
0.90	86%	77%
1.0	100%	100%

### 7.2.5 Obtención de genotipos por GT-seq

En la secuenciación del panel para GT-seq con 118 SNPs se obtuvo una profundidad de lectura promedio de 31.7x. Sin embargo, 21 SNPs presentaron baja profundidad de lectura (<6x; Fig. 16), un SNP no amplificó (SNP 4048\_586\_118) y otro SNP se descartó por un bajo porcentaje de genotipificación en todas las muestras utilizadas (2.1% de las muestras; SNP 9717\_50\_71).

Por otra parte, dos SNPs fueron descartados por encontrarse en la secuencia mitocondrial de la especie (SNP 11208\_39\_117 y SNP 9814\_36\_52); esto se confirmó con un análisis de blast (datos no mostrados) al genoma mitocondrial (n° de acceso GenBank NC\_009626.1). Estos dos SNPs estuvieron sobrerrepresentados con el 23% de las lecturas totales objetivo (Fig. 16).

En cuanto al SNP de identificación de sexo, el 100% de los organismos que estaban identificados visualmente (presencia de petasma en machos) coincidieron con el genotipo esperado (Hembras C/T, machos C/C).

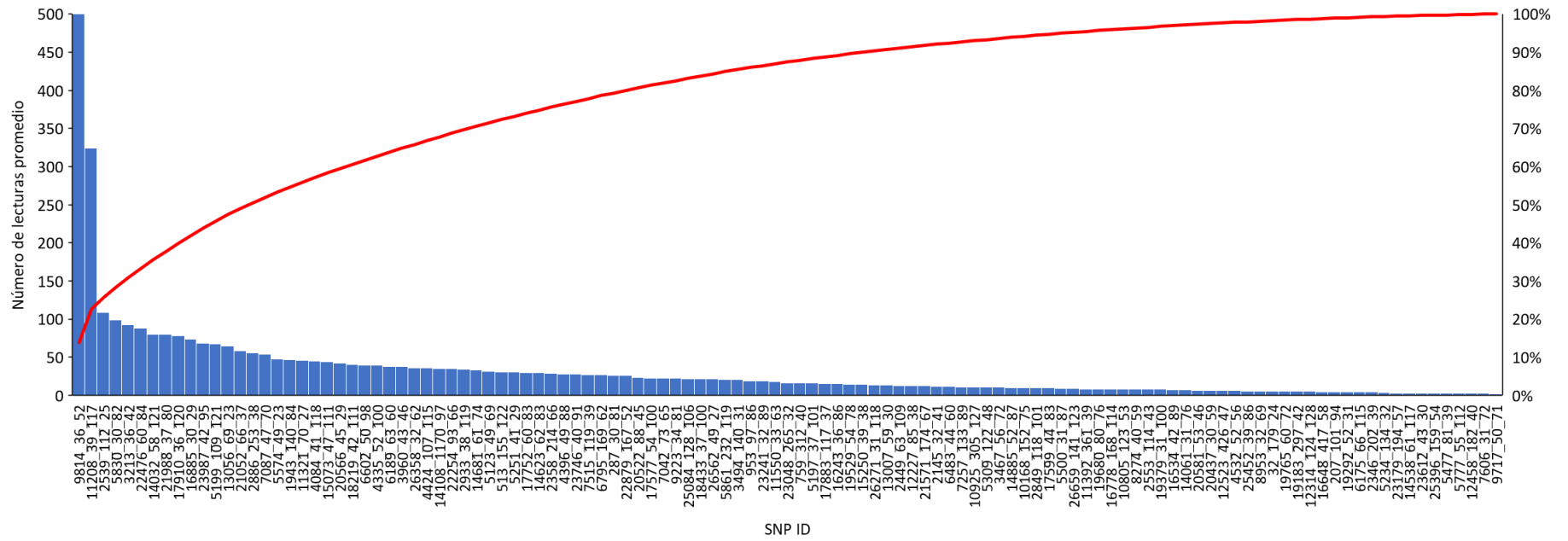


Figura 16. Promedio de lecturas objetivo para cada locus del panel de 118 SNPs.

### 7.2.6 Validación de panel de SNPs para análisis de parentesco

Para los análisis de parentesco se utilizaron 20 familias de camarón ( $n=120$ ; 20 hembras, 20 machos, 80 progenie) ya que 10 individuos fueron descartados de los análisis por bajo porcentaje de genotipificación ( $<50\%$ ), además se realizaron los análisis con 91 SNPs (22 SNPs se descartaron por bajo porcentaje de genotipos).

En los resultados con Cervus, se obtuvo una probabilidad combinada de no-exclusión de  $3 \times 10^{-8}$  para la asignación de la progenie a ambos padres. Los porcentajes de asignación de la progenie a ambos padres fue alto con Cervus (78-99%), aunque no alcanzaron el 100% a diferencia del análisis con Colony, con el cual el 100% de la progenie fue asignada a ambos padres (Tabla 8). Cabe mencionar que los individuos que no fueron asignados con el análisis en Cervus, pertenecieron principalmente a la misma familia. Todas las asignaciones obtenidas coincidieron con las previamente reportadas con un panel comercial en un análisis independiente.

En el análisis con Colony, la probabilidad de las asignaciones obtenidas para ambos padres fue alta (0.99-1.00). Para la conformación de la progenie en familias, 20 familias fueron inferidas correctamente con una probabilidad de inclusión (probabilidad de que los miembros de la familia hayan sido incluidos) de 0.9988-1.0 y una probabilidad de exclusión de 0.1708-1.00.

**Tabla 8.** Porcentajes de asignación de progenie a ambos padres ( $n=80$ )

Proporción de padres muestreados	Cervus		Colony
	Nivel de confianza		
	90%	95%	
0.90	94%	78%	100%
1.0	99%	91%	100%

## 8. DISCUSIÓN

### 8.1 Panel de SNPs con 2bRAD

El número de SNPs obtenido con 2bRAD fue mayor al esperado considerando los resultados del análisis *in silico*. Con 2 bases selectivas (19,105 SNPs) fue 2.5 veces mayor al calculado en el análisis *in silico* (7,669) y 6.9 veces mayor con 4 bases selectivas (2,874 SNPs vs 418 SNPs en el análisis *in silico*). Esto podría deberse a que las secuencias de la base de datos pública (Perez-Enriquez *et al.*, 2018c) utilizada en el análisis *in silico* para estimar la proporción de sitios variables fueron obtenidas por NextRAD, el cual es un método distinto que utiliza un transposoma de Nextera que fragmenta y etiqueta el ADN.

Además, el porcentaje de alineamiento de las secuencias al genoma de referencia (63.4-72%) puede ser optimizado debido a que el genoma de referencia disponible cuando se realizó el análisis cubre sólo del 64-68% del genoma estimado para la especie (1.66 Gb de los 2.45 y 2.6 Gb estimados con citometría de flujo y análisis de k-mer, respectivamente) (Zhang *et al.*, 2019), por lo que, utilizar los genomas de referencia recientemente publicados (Peng *et al.*, 2023; Perez-Enriquez *et al.*, 2024a), los cuales están más completos (1.87 y 2.055 Gb, respectivamente) incrementaría el porcentaje de alineamiento y por lo tanto, el número de SNPs del panel.

La densidad de SNPs obtenida con dos bases selectivas fue similar a otros trabajos en *P. vannamei* con el mismo método (23k SNPs), pero sin utilizar el genoma de referencia de Zhang *et al.* (2019) (Wang *et al.*, 2017b; Wang *et al.*, 2019a; Wang *et al.*, 2019b). Aunque en estos estudios, se asume que no utilizan bases selectivas en los adaptadores, ya que no lo especifican, al parecer usan la misma enzima de restricción (Bcgl).

El método de 2bRAD puede ser una opción viable para estudios que requieran paneles de SNPs de mediana y alta densidad (miles de SNPs), como GWAS (Wang *et al.*, 2019b; Yu *et al.*, 2019; Medrano-Mendoza *et al.*, 2023), selección genómica (Wang *et al.*, 2017b; Wang *et al.*, 2019a), mapas de ligamiento (Peng *et al.*, 2020), entre otros. Sin embargo, para otro tipo de estudios, por ejemplo, de parentesco (Perez-Enriquez y Max-Aguilar, 2016), es posible utilizar una menor densidad de SNPs, principalmente si se requiere optimizar el número de individuos a analizar en



una misma línea de secuenciación y reducir con esto el costo del análisis; por lo que otras opciones como GT-seq, pueden ser viables cuando no se requiere una alta densidad de marcadores.

## **8.2 Panel de 118 SNPs y obtención de genotipos con GT-seq**

Utilizando una base de datos de secuenciación pública (proyecto NCBI-SRA PRJNA492152; Pérez-Enriquez *et al.*, 2018c), se desarrolló un panel de 118 SNPs, el cual incluye SNPs enfocados a estructura genética y asignación poblacional, análisis de parentesco e identificación de sexo; este panel fue adaptado para la obtención de genotipos en una sola reacción multiplex por individuo con el método de GT-seq (Campbell *et al.*, 2015), lo cual reduce el tiempo de preparación de librerías. Además, el objetivo de enfocarse en un panel de baja densidad permite incluir un mayor número de muestras en una sola línea de secuenciación, lo que a su vez reduce los costos de genotipificación; por ejemplo, las celdas disponibles para un secuenciador Miseq (Illumina) tienen una capacidad de 1, 4, 15 y 25 millones de lecturas, en las que sería posible analizar 175, 700, 2630 y 4385 individuos respectivamente (20% PhiX, 40× de profundidad de lectura, 114 SNPs).

Dentro del panel de 118 SNPs, se observó una sobrerrepresentación de las lecturas de los SNPs 9814\_36\_52 y 11208\_39\_117 (23% del total de lecturas objetivo), los cuales resultaron ubicarse en el ADN mitocondrial. El problema de la sobrerrepresentación de pocos SNPs es que afecta la profundidad de lectura de los demás loci, lo que a su vez provoca la pérdida de genotipos; se observó baja profundidad de lectura <6× en 21 SNPs (lo que ocasionó el descarte de SNPs en los análisis posteriores de asignación poblacional y parentesco). El problema de la sobrerrepresentación en GT-seq ha sido reportada en otros estudios (Baetscher *et al.*, 2018; Bootsma *et al.*, 2020); en Bootsma *et al.* (2020) realizaron tres rondas de prueba en las que eliminan los SNPs con alto grado de amplificación (identificando la proporción de lecturas en el 10% de los loci con mayor número de lecturas) y con esto mejoraron la homogenización de lecturas entre loci. Considerando lo anterior, la eliminación de estos dos SNPs mejorará la genotipificación del panel con GT-seq.

Para el uso de este panel en futuros análisis, se recomienda incluir sólo 114 SNPs durante la preparación de librerías, descartando los dos SNPs anteriormente mencionados, así como los SNPs 4048\_586\_118 y 9717\_50\_71, por su deficiente amplificación.

### **8.3 Caracterización genómica de poblaciones de cultivo de *P. vannamei* en México y validación de panel de SNPs de baja densidad para análisis de asignación poblacional y trazabilidad**

En el análisis de Structure, al comparar el origen de los lotes de cultivo que fueron reportados por los laboratorios de producción, y los resultados obtenidos para K=6 (Fig. 12; Tabla 6) se logró inferir exitosamente el origen algunos lotes (como el 5 y el 55), y distinguir claramente los organismos del lote con origen reportado en México, con los lotes de importación, esto último, acorde también con el DAPC (Fig. 15). Por otra parte, se observan lotes reportados con el mismo origen, en grupos genéticos distintos, esto principalmente para los lotes de cultivo importados de Ecuador; para el caso de los lotes con origen en Texas, se distinguen dos grupos genéticos: los lotes 5 y 55 (Grupo 3; Tabla 6) y los lotes 2 y 13 en el grupo genético 6 (Tabla 6). Esto puede deberse a que provienen de distintos lotes de origen y/o distintos eventos de importación; o que, aunque hayan provenido de un mismo lote importado, el manejo y reproducción entre los laboratorios de producción es distinto; por ejemplo, algunos laboratorios mantuvieron pocas familias (Rafael Ruíz, comunicación personal), lo que en un par de generaciones ocasionaría cambios importantes en las frecuencias alélicas.

En la validación del panel de SNPs para trazabilidad, casi la mitad de las muestras (46%; n=74) fueron asignadas a las poblaciones de referencia de origen conocido reportadas por Perez-Enriquez *et al.* (2018c); de las asignaciones obtenidas 18 individuos fueron asignados a un origen que no corresponde al reportado por el laboratorio de producción y otros 9 individuos del lote 42 (origen reportado como Buenavista) fueron asignados a la población de referencia de Ecuador. Hay que considerar, que la falta de asignación de muestras de Texas y Ecuador puede deberse, como ya se mencionó en el párrafo anterior, a diferencias en el manejo entre los laboratorios de producción y en cuanto al lote 4 (origen reportado Hawái), no se contó con una población de referencia, por lo que no se obtuvo asignación poblacional de esas muestras, pero en el DAPC (Fig. 15B) los organismos de este lote se separan claramente de los demás. Otra consideración es que no se observa una clara diferencia genética entre algunos orígenes de referencia; se observa

que las muestras de referencia de Centroamérica y Ecuador se encuentran en un mismo grupo genético (Fig. 12).

Otros estudios, con especies pesqueras, obtuvieron altos porcentajes de asignación a su población de origen; 98% de asignación en merluza *Merluccius merluccius* (Nielsen *et al.*, 2012), de 13-100% de asignación en salmón *Oncorhynchus tshawytscha* (Clemento *et al.*, 2014), 81-83% en atún aleta azul *Thunnus thynnus* (Rodríguez-Ezpeleta *et al.*, 2019). Así como bajos porcentajes de asignaciones incorrectas, 1.05% [comparación entre marcaje físico y asignación con panel de SNPs; Clemento *et al.* (2014)], 2% y 10% (Rodríguez-Ezpeleta *et al.*, 2019). Aunque, el número de SNPs utilizado es similar al de este trabajo; por ejemplo, en Clemento *et al.* (2014), se utilizó un mayor número de muestras como población de referencia (n=8,000).

Ante la falta de una base de datos de referencia mayor, en situaciones en donde existen lotes de camarón incautados, una estrategia que puede ser implementada con este panel de SNPs, es analizar las muestras del producto incautado y tomar muestras del presunto origen del producto, para comparar los perfiles genéticos, además de incluir perfiles genéticos de otras muestras independientes como referencia.

Por otra parte, Hayes *et al.* (2005), para el cultivo de Salmón del Atlántico en Noruega, evalúan *in silico* tres estrategias de trazabilidad, considerando el número de marcadores genéticos y el número de reproductores a analizar por granja, en donde consideran que la estrategia menos costosa podría ser genotipificar 200 SNPs y 100 peces de cada núcleo en unidad de producción. Un punto importante que considerar es la organización de cada industria; por ejemplo, si está organizada en tres niveles, donde los individuos en venta para consumo del cliente son principalmente los nietos de los reproductores del núcleo. Sin embargo, en industrias como la de camarón en México, la organización es distinta y este tipo de estrategias son logísticamente difíciles de implementar.

Dentro de las recomendaciones para futuros proyectos de investigación, se encuentra la implementación en México de proyectos como FishPopTrace, el cual es un proyecto internacional en Europa (<http://fishpoptrace.jrc.ec.europa.eu/>) enfocado, entre otras cosas, al desarrollo e

implementación herramientas tecnológicas de trazabilidad de especies pesqueras y sus productos, los cuales permitan mantener una amplia base de datos. A la fecha, no existen otros estudios enfocados a la posible trazabilidad de *P. vannamei*, tanto en cultivo como en muestras silvestres.

Por otra parte, un resultado relevante que se pudo observar en los análisis de estructura genética (tanto con Structure como con DAPC), es que no hay una separación entre las poblaciones silvestres y algunos lotes de cultivo; por ejemplo en la gráfica de bandera para K=7 (Fig. 14), se pueden observar perfiles genéticos similares entre las muestras silvestres y los lotes 11, 34, 39 y 42. En un estudio previo con microsatélites, en muestras de 2012-2013, se reportó un estimado de 7.1% de posibles escapes de camarones de cultivo al medio silvestre, y se observa una diferenciación genética entre poblaciones silvestres y de cultivo (Perez-Enriquez *et al.*, 2018a). Sin embargo, en estudios más recientes se reportan resultados similares a este trabajo, en donde con un panel de SNPs comercial, se observan perfiles genéticos similares entre lotes de cultivo y poblaciones silvestres (Ruiz-Luna *et al.*, 2023; Perez-Enriquez *et al.*, 2024b). Además, Ruiz-Luna *et al.* (2023) reportan la liberación de organismos de cultivo a lagunas costeras de Sinaloa para repoblamiento en 2019-2020; en estas prácticas, es importante que los organismos que se utilizan para repoblamiento tengan perfiles genéticos similares a las poblaciones silvestres (Miller y Kapuscinski, 2003; Perez-Enriquez *et al.*, 2013) y evitar modificar a largo plazo la composición genética de las poblaciones silvestres, lo cual por ejemplo, puede disminuir la variabilidad genética y su respuesta ante cambios ambientales (Weir y Grant, 2005; Danancher y Garcia-Vazquez, 2011; Glover *et al.*, 2013; Zhang *et al.*, 2013). Aunado a estas prácticas, el escape accidental de organismos de cultivo al medio silvestre puede ocurrir, por ejemplo, durante la temporada de huracanes. Un estudio más amplio para evaluar la posible introgresión de organismos de cultivo al medio silvestre, incluyendo muestras históricas silvestres es recomendable.

#### **8.4 Validación de panel de SNPs de baja densidad para análisis de parentesco en cultivo de camarón blanco *Penaeus vannamei***

En la validación del panel de SNPs de baja densidad, se obtuvo que el 100% de la progenie fue asignada a ambos padres correctamente, así como la asignación de las 20 familias esperadas con

Colony. Sin embargo, en el análisis con el programa Cervus, aunque se obtuvo un alto porcentaje de asignación (78-99%), no se alcanzó el 100%, lo cual pudo deberse a perfiles genéticos incompletos por baja profundidad de lectura de algunos SNPs (como se menciona en la sección 8.2 y su posible solución).

La robustez de un panel de SNPs en análisis de parentesco puede ser afectada por el número de marcadores, el MAF (frecuencia alélica del alelo menor) de esos marcadores, genotipos faltantes y el error de genotipificación (Anderson y Garza, 2006; Liu *et al.*, 2016; Premachandra *et al.*, 2019; Massault *et al.*, 2021). El número de SNPs utilizado fue similar a los reportados para otras especies acuícolas como el salmón del Atlántico con 94 SNPs (Holman *et al.*, 2017), trucha arcoíris con 95 SNPs (Liu *et al.*, 2016), el panel *VannaPlus*<sup>®</sup> recientemente publicado para *P. vannamei* con 96 SNPs (Lisboa Silva *et al.*, 2022); Anderson y Garza (2006) mencionan que de 60-100 SNPs pueden ser suficientes para obtener el 100% de asignación en análisis de pedigrí. En un estudio previo en *P. vannamei* con un panel comercial de 76 SNPs, se obtuvieron porcentajes de asignación del 94-96%, incluso disminuyendo a 50 SNPs con un MAF de 0.3 (Perez-Enriquez y Max-Aguilar, 2016). Los SNPs con mayor MAF son más robustos que aquellos con menor MAF en los análisis de asignación de parentesco (Anderson y Garza, 2006; Perez-Enriquez y Max-Aguilar, 2016), por lo que, para el desarrollo del panel se seleccionaron SNPs con MAF >0.2, lo cual resultó adecuado. En cuanto al error de genotipificación, aunque no se incluyó en este trabajo un cálculo de error de genotipos, bajos porcentajes de error han sido reportado para SNPs utilizando GT-seq, por ejemplo, 0.50% (Schmidt *et al.*, 2020), 0.54% (Shedd *et al.*, 2022) ó 6.8% y 10.7% cuando se utiliza ADN de baja calidad de muestras fecales (Hayward *et al.*, 2022). Además, en la comparación de los perfiles genéticos de padres e hijos de las asignaciones obtenidas, no se presentaron en su mayoría discordancias entre sus genotipos (datos no mostrados).

Por otra parte, cuando se compararon los resultados de la validación *in silico* con la validación realizada con muestras de familias de camarón, se observó una disminución importante en la probabilidad de no-exclusión ( $5.87 \times 10^{-15}$  vs  $3 \times 10^{-8}$ ); si consideramos lo ya mencionado a lo largo de la discusión, asumimos que se debió a los genotipos faltantes observados en los datos reales; esto ya ha sido mencionado en otro trabajo (Vandeputte *et al.*, 2011), por lo que debe ser un

factor a considerar en futuros trabajos donde partimos de análisis *in silico* para el desarrollo de paneles de marcadores.

Siendo que la meta óptima de un panel de SNPs para análisis de pedigrí es obtener el 100% de asignación de los organismos con alto grado certidumbre, al menor costo posible, el panel desarrollado en este trabajo cumple con gran parte del objetivo; el costo aproximado de genotipificación es de 10.13 USD para un mínimo de 192 muestras, y de 8.45 USD para 768 muestras sin incluir la extracción de ADN (Anexo I). El que sea en un panel de SNPs de baja densidad y con la optimización realizada en la genotipificación, lo convierte en una opción asequible, aunado al hecho de evitar el envío de muestras al extranjero, ya que se puede realizar el análisis en un laboratorio de genética en México, como el CIBNOR.

## 9. CONCLUSIONES

La caracterización genómica de lotes de cultivo permitió distinguir hasta seis grupos genéticos distintos con diferentes orígenes de importación en México, además, de una clara diferenciación genética entre el lote de cultivo con origen en México y los lotes con origen de importación; en la validación del panel de SNPs de baja densidad para trazabilidad, se logró la asignación del 46% de las muestras “ciegas” a los cinco grupos genéticos de origen conocido previamente reportados para México. La falta de asignación del resto de los individuos, se debió a distintas razones, entre las cuales fue la ausencia de muestras de referencia para algunos orígenes (como en el caso del lote de importación de Hawái), así como los cambios en las frecuencias alélicas debido a diferentes tipos de manejo dentro y entre los laboratorios de producción, lo que dificulta este tipo de análisis, sin embargo este panel puede ser utilizado para verificar el origen declarado de un lote de cultivo con la presunta población de origen.

En la caracterización genómica de lotes de cultivo y muestras silvestres, se observó similitud en los perfiles genéticos entre algunos lotes de cultivo y las muestras silvestres, lo que indicaría la posible introgresión de organismos de cultivo en las poblaciones silvestres.

El panel de SNPs de baja densidad desarrollado en este trabajo fue validado también para análisis de parentesco con hasta el 100% de asignación de la progenie a ambos padres, así como con la asignación correcta de las familias esperadas; esto aunado a la estandarización de genotipificación enfocada a reducir el tiempo y costo de análisis, lo convierte en una buena opción para los productores de camarón que requieran estos análisis.

El panel de baja densidad cumple con tres objetivos en una sola reacción multiplex, con una misma plataforma de obtención de genotipos (GT-seq): 1) estudiar la estructura genética y la asignación de muestras a una población de referencia, 2) análisis de parentesco (asignación de progenie a ambos padres y asignación de familias) y 3) identificación de sexo.

Por otra parte, el panel de SNPs desarrollado con 2bRAD es una herramienta para futuros análisis que requieran alta densidad de marcadores moleculares, como estudios de asociación a lo largo

del genoma (GWAS) y Selección Genómica en *P. vannamei*, los cuales permitirían estudiar la arquitectura genética de caracteres económicos de interés, así como optimizar la selección de dichas características.



## 10. LITERATURA CITADA

- Ababouch, L., Gandini, G. y Ryder, J. (2005). *Causes of detentions and rejections in international fish trade*. FAO, Rome.
- Anderson, E.C. y Garza, J.C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172(4): 2567-2582. doi: 10.1534/genetics.105.048074
- Baetscher, D.S., Clemento, A.J., Ng, T.C., Anderson, E.C. y Garza, J.C. (2018). Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources*, 18(2): 296-305. doi: 10.1111/1755-0998.12737
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. y Johnson, E.A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10): e3376. doi: 10.1371/journal.pone.0003376
- Barbanti, A., Torrado, H., Macpherson, E., Bargelloni, L., Franch, R., Carreras, C. y Pascual, M. (2020). Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. *Molecular Ecology Resources*, 20(3): 795-806. doi: 10.1111/1755-0998.13144
- Bolger, A.M., Lohse, M. y Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114-2120. doi: 10.1093/bioinformatics/btu170
- Bootsma, M.L., Gruenthal, K.M., McKinney, G.J., Simmons, L., Miller, L., Sass, G.G. y Larson, W.A. (2020). A GT-seq panel for walleye (*Sander vitreus*) provides important insights for efficient development and implementation of amplicon panels in non-model organisms. *Molecular Ecology Resources*, 20(6): 1706-1722. doi: 10.1111/1755-0998.13226
- Brown, C., Miltiadou, D. y Tsigenopoulos, C.S. (2015). Prevalence and survival of escaped European seabass *Dicentrarchus labrax* in Cyprus identified using genetic markers. *Aquaculture Environment Interactions*, 7: 49-59. doi: 10.3354/aei00135
- Brown, S.S., Chen, Y.-W., Wang, M., Clipson, A., Ochoa, E. y Du, M.-Q. (2017). PrimerPooler: Automated primer pooling to prepare library for targeted sequencing. *Biology Methods and Protocols*, 2: 1-10. doi: 10.1093/biomethods/bpx006
- Bylemans, J., Maes, G.E., Diopere, E., Cariani, A., Senn, H., Taylor, M.I., Helyar, S., Bargelloni, L., Bonaldo, A., Carvalho, G., Guarniero, I., Komen, H., Martinsohn, J.T., Nielsen, E.E., Tinti, F., Volckaert, F.A.M. y Ogden, R. (2016). Evaluating genetic traceability methods for captive-bred marine fish and their applications in fisheries management and wildlife forensics. *Aquaculture Environment Interactions*, 8: 131-145. doi: 10.3354/aei00164
- Campbell, N.R., Harmon, S.A. y Narum, S.R. (2015). Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4): 855-867. doi: 10.1111/1755-0998.12357
- Castillo-Juárez, H., Campos-Montes, G.R., Caballero-Zamora, A. y Montaldo, H.H. (2015). Genetic improvement of Pacific white shrimp [*Penaeus (Litopenaeus) vannamei*]: perspectives for genomic selection. *Frontiers in Genetics*, 6: 93. doi: 10.3389/fgene.2015.00093
- Chauhan, T. y Rajiv, K. (2010). Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology*, 01(04): 281-291. doi: 10.4236/abb.2010.14037
- Chen, B., Zhong, P., Wu, X., Peng, K., Sun, Y., Chen, X., Zhao, H., Xu, Z., Liu, J., Li, H., Li, C. y Huang, W. (2022). Construction of a genetic linkage map, QTLs mapping for low salinity and growth-related traits and identification of the candidate genes in Pacific white shrimp (*Litopenaeus vannamei*). *Aquaculture Reports*, 22: 100978. doi: 10.1016/j.aqrep.2021.100978

- Clemento, A.J., Crandall, E.D., Garza, J.C. y Anderson, E.C. (2014). Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. *Fishery Bulletin*, 112(2-3): 112-130. doi: 10.7755/FB.112.2-3.2
- Collins, F.S., Brooks, L.D. y Chakravarti, A. (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*, 8(12): 1229-1231. doi: 10.1101/gr.8.12.1229
- CONAPESCA. (2021). *Anuario estadístico de acuacultura y pesca 2021*. Disponible en: <https://www.gob.mx/conapesca/documentos/anuario-estadistico-de-acuacultura-y-pesca> [consulta: 20 febrero 2024]
- CONAPESCA. (2017). *Anuario estadístico de acuacultura y pesca 2017*. Disponible en: <https://www.gob.mx/conapesca/documentos/anuario-estadistico-de-acuacultura-y-pesca> [consulta: 16 noviembre 2020]
- Cruz, P., Mejia-Ruiz, C.H., Perez-Enriquez, R. y Ibarra, A.M. (2002). Isolation and characterization of microsatellites in Pacific white shrimp *Penaeus (Litopenaeus) vannamei*. *Molecular Ecology Notes*, 2(3): 239-241. doi: 10.1046/j.1471-8286.2002.00211.x
- Dai, P., Kong, J., Liu, J., Lu, X., Sui, J., Meng, X. y Luan, S. (2020). Evaluation of the utility of genomic information to improve genetic evaluation of feed efficiency traits of the Pacific white shrimp *Litopenaeus vannamei*. *Aquaculture*, 527: 735421. doi: 10.1016/j.aquaculture.2020.735421
- Danancher, D. y Garcia-Vazquez, E. (2011). Genetic population structure in flatfishes and potential impact of aquaculture and stock enhancement on wild populations in Europe. *Reviews in Fish Biology and Fisheries*, 21(3): 441-462. doi: 10.1007/s11160-011-9198-6
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. y 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15): 2156-2158. doi: 10.1093/bioinformatics/btr330
- Davis, G.P. y Hetzel, D.J.S. (2000). Integrating molecular genetic technology with traditional approaches for genetic improvement in aquaculture species. *Aquaculture Research*, 31(1): 3-10. doi: 10.1046/j.1365-2109.2000.00438.x
- Dong, S., Kong, J., Zhang, T., Meng, X. y Wang, R. (2006). Parentage determination of Chinese shrimp (*Fenneropenaeus chinensis*) based on microsatellite DNA markers. *Aquaculture*, 258: 283-288. doi: 10.1016/j.aquaculture.2006.04.044
- Doyle, R.W., Perez-Enriquez, R., Takagi, M. y Taniguchi, N. (2001). Selective recovery of founder genetic diversity in aquacultural broodstocks and captive, endangered fish populations. *Genetica*, 111: 291-304. doi: <https://doi.org/10.1023/A:1013772205330>
- Dupuis, J.R., Bremer, F.T., Kauwe, A., San Jose, M., Leblanc, L., Rubinoff, D. y Geib, S.M. (2018). HiMAP: Robust phylogenomics from highly multiplexed amplicon sequencing. *Molecular Ecology Resources*, 18(5): 1000-1019. doi: 10.1111/1755-0998.12783
- Earl, D.A. y VonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2): 359-361. doi: 10.1007/s12686-011-9548-7
- Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, 16(12): 551-558. doi: 10.1016/S0168-9525(00)02139-9
- Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P. y Guyomard, R. (1998). Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow

- trout (*Oncorhynchus mykiss*) hatchery populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(3): 715-725. doi: 10.1139/cjfas-55-3-715
- Excoffier, L. y Lischer, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3): 564-567. doi: 10.1111/j.1755-0998.2010.02847.x
- FAO (2024). *Fishery and Aquaculture Statistics – Yearbook 2021*. FAO, Rome. doi: <https://doi.org/10.4060/cc9523en>
- FAO (2006). *Programa de información de especies acuáticas. Penaeus vannamei*. Disponible en: [https://www.fao.org/figis/pdf/fishery/culturedspecies/Penaeus\\_vannamei/es?title=FAO%20Fisheries%20%26%20Aquaculture%20-%20Programa%20de%20informaci%F3n%20de%20especies%20acu%E1ticas%20-%20Penaeus%20vannamei%20%28Boone%2C%201931%29](https://www.fao.org/figis/pdf/fishery/culturedspecies/Penaeus_vannamei/es?title=FAO%20Fisheries%20%26%20Aquaculture%20-%20Programa%20de%20informaci%F3n%20de%20especies%20acu%E1ticas%20-%20Penaeus%20vannamei%20%28Boone%2C%201931%29) [consulta: 16 noviembre 2020]
- FAO (2020). *Fishery and aquaculture statistics. Global aquaculture production 1950-2018 (FishstatJ)*. Disponible en: [www.fao.org/fishery/statistics/software/fishstatj/en](http://www.fao.org/fishery/statistics/software/fishstatj/en) [consulta: 10 enero 2021]
- Faust, E., Halvorsen, K.T., Andersen, P., Knutsen, H. y André, C. (2018). Cleaner fish escape salmon farms and hybridize with local wrasse populations. *Royal Society Open Science*, 5(3): 171752. doi: 10.1098/rsos.171752
- Freitas, J., Vaz-Pires, P. y Câmara, J.S. (2020). From aquaculture production to consumption: Freshness, safety, traceability and authentication, the four pillars of quality. *Aquaculture*, 518: 734857. doi: 10.1016/j.aquaculture.2019.734857
- García, D.K. y Alcivar-Warren, A. (2007). Characterization of 35 new microsatellite genetic markers for the Pacific whiteleg shrimp, *Litopenaeus vannamei*: Their usefulness for studying genetic diversity of wild and cultured stocks, tracing pedigree in breeding programs, and linkage mapping. *Journal of Shellfish Research*, 26(4): 1203-1216. doi: 10.2983/0730-8000(2007)26[1203:CONMGM]2.0.CO;2
- Gjedrem, T. y Baranski, M. (2009). *Selective breeding in aquaculture: An Introduction*. Springer Netherlands, Dordrecht. doi: 10.1007/978-90-481-2773-3
- Gjedrem, T., Robinson, N. y Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture*, 350-353: 117-129. doi: 10.1016/j.aquaculture.2012.04.008
- Gjerde, B., Gjøen, H.M. y Villanueva, B. (1996). Optimum designs for fish breeding programmes with constrained inbreeding mass selection for a normally distributed trait. *Livestock Production Science*, 47(1): 59-72. doi: 10.1016/S0301-6226(96)01000-7
- Gjerde, B. y Rye, M. (1998). Design of breeding programmes in aquaculture species: Possibilities and constraints. En: *Genetics and breeding of Mediterranean aquaculture species*. CIHEAM, Zaragoza, pp. 181-192
- Glover, K.A. (2008). Genetic characterisation of farmed rainbow trout in Norway: intra- and inter-strain variation reveals potential for identification of escapees. *BMC Genetics*, 9(1): 87. doi: 10.1186/1471-2156-9-87
- Glover, K.A. (2010). Forensic identification of fish farm escapees: The Norwegian experience. *Aquaculture Environment Interactions*, 1: 1-10. doi: 10.3354/aei00002
- Glover, K.A., Hansen, M.M. y Skaala, Ø. (2009). Identifying the source of farmed escaped Atlantic salmon (*Salmo salar*): Bayesian clustering analysis increases accuracy of assignment. *Aquaculture*, 290(1-2): 37-46. doi: 10.1016/j.aquaculture.2009.01.034

- Glover, K.A., Pertoldi, C., Besnier, F., Wennevik, V., Kent, M. y Skaala, Ø. (2013). Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC Genetics*, 14: 74. doi: 10.1186/1471-2156-14-74
- Glover, K.A., Skilbrei, O.T. y Skaala, Ø. (2008). Genetic assignment identifies farm of origin for Atlantic salmon *Salmo salar* escapees in a Norwegian fjord. *ICES Journal of Marine Science*, 65(6): 912-920. doi: 10.1093/icesjms/fsn056
- Han, J.E., Tang, K.F.J., Tran, L.H. y Lightner, D.V. (2015). *Photorhabdus* insect-related (Pir) toxin-like genes in a plasmid of *Vibrio parahaemolyticus*, the causative agent of acute hepatopancreatic necrosis disease (AHPND) of shrimp. *Diseases of Aquatic Organisms*, 113: 33-40. doi: 10.3354/dao02830
- Harney, E., Lachambre, S., Roussel, S., Huchette, S., Enez, F., Morvezen, R., Haffray, P. y Boudry, P. (2018). Transcriptome based SNP discovery and validation for parentage assignment in hatchery progeny of the European abalone *Haliotis tuberculata*. *Aquaculture*, 491: 105-113. doi: 10.1016/j.aquaculture.2018.03.006
- Hastein, T., Hill, B.J., Berthe, F. y Lightner, D.V. (2001). Traceability of aquatic animals. *Revue Scientifique et Technique de l'OIE*, 20(2): 564-583. doi: 10.20506/rst.20.2.1300
- Hayes, B., Sonesson, A.K. y Gjerde, B. (2005). Evaluation of three strategies using DNA markers for traceability in aquaculture species. *Aquaculture*, 250: 70-81. doi: 10.1016/j.aquaculture.2005.03.008
- Hayward, K.M., Clemente-Carvalho, R.B.G., Jensen, E.L., de Groot, P.V.C., Branigan, M., Dyck, M., Tschritter, C., Sun, Z. y Lougheed, S.C. (2022). Genotyping-in-thousands by sequencing (GT-seq) of noninvasive faecal and degraded samples: A new panel to enable ongoing monitoring of Canadian polar bear populations. *Molecular Ecology Resources*, 22(5): 1906-1918. doi: 10.1111/1755-0998.13583
- Herbinger, C.M., Doyle, R.W., Pitman, E.R., Paquet, D., Mesa, K.A., Morris, D.B., Wright, J.M. y Cook, D. (1995). DNA fingerprint based analysis of paternal and maternal effects on offspring growth and survival in communally reared rainbow trout. *Aquaculture*, 137(1-4): 245-256. doi: [https://doi.org/10.1016/0044-8486\(95\)01109-9](https://doi.org/10.1016/0044-8486(95)01109-9)
- Holman, L.E., Garcia de la serrana, D., Onoufriou, A., Hillestad, B. y Johnston, I.A. (2017). A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture*, 476: 59-64. doi: 10.1016/j.aquaculture.2017.04.001
- Houston, R.D., Taggart, J.B., Cézard, T., Bekaert, M., Lowe, N.R., Downing, A., Talbot, R., Bishop, S.C., Archibald, A.L., Bron, J.E., Penman, D.J., Davassi, A., Brew, F., Tinch, A.E., Gharbi, K. y Hamilton, A. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics*, 15: 90. doi: 10.1186/1471-2164-15-90
- Jerry, D.R., Evans, B.S., Kenway, M. y Wilson, K. (2006). Development of a microsatellite DNA parentage marker suite for black tiger shrimp *Penaeus monodon*. *Aquaculture*, 255(1-4): 542-547. doi: 10.1016/j.aquaculture.2005.11.026
- Jerry, D.R., Preston, N.P., Crocos, P.J., Keys, S., Meadows, J.R.S. y Li, Y. (2004). Parentage determination of Kuruma shrimp *Penaeus (Marsupenaeus) japonicus* using microsatellite markers (Bate). *Aquaculture*, 235(1-4): 237-247. doi: 10.1016/j.aquaculture.2004.01.019
- Jombart, T. (2008). Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11): 1403-1405. doi: 10.1093/bioinformatics/btn129
- Jones, D.B., Nguyen, H.T., Khatkar, M.S., Simma, D.B., Jerry, D.R., Raadsma, H.W., van der Steen, H., Prochaska, J. y Zenger, K.R. (2020). The identification of a major sex QTL in the white-leg

- shrimp, *Litopenaeus vannamei*. *Aquaculture*, 529: 735673. doi: 10.1016/j.aquaculture.2020.735673
- Jones, O.R. y Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, 10(3): 551-555. doi: 10.1111/j.1755-0998.2009.02787.x
- Kalinowski, S.T., Taper, M.L. y Marshall, T.C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, 16(5): 1099-1106. doi: 10.1111/j.1365-294X.2007.03089.x
- Karlsson, S., Diserud, O.H., Fiske, P. y Hindar, K. (2016). Widespread genetic introgression of escaped farmed Atlantic salmon in wild salmon populations. *ICES Journal of Marine Science*, 73(10): 2488-2498. doi: 10.1093/icesjms/fsw121
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. y Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5): 1179-1191. doi: 10.1111/1755-0998.12387
- Lee, C.T., Chen, I.T., Yang, Y.T., Ko, T.P., Huang, Y.T., Ko, T.P., Huang, Y.T., Huang, J.Y., Huang, M.F., Lin, S.J., Chen, C.Y., Lin, S.S., Lightner, D.V., Wang, H.C., Wang, A.H.J., Wang, H.C., Hor, L.I. y Lo, C.F. (2015). The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences USA*, 112(39): 10798-10803. doi: 10.1073/pnas.1517100112
- Li, H. y Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14): 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. y 1000 Genomes Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16): 2078-2079. doi: 10.1093/bioinformatics/btp352
- Lisboa Silva, N.M., Ianella, P., Beleza Yamagishi, M.E., Rocha, J.L., Teixeira, A.K., Galvão Farias, F., Guerrelhas, A.C. y Rodrigues Caetano, A. (2022). Development and validation of a low-density SNP panel for paternity and kinship analysis and evaluation of genetic variability and structure of commercial Pacific white shrimp (*Litopenaeus vannamei*) populations from Brazil. *Aquaculture*, 560: 738540. doi: 10.1016/j.aquaculture.2022.738540
- Lischer, H.E.L. y Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2): 298-299. doi: 10.1093/bioinformatics/btr642
- Liu, S., Palti, Y., Gao, G. y Rexroad, C.E. (2016). Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture*, 452: 178-182. doi: 10.1016/j.aquaculture.2015.11.001
- Liu, T., Li, Q., Kong, L. y Yu, H. (2017). Comparison of microsatellites and SNPs for pedigree analysis in the Pacific oyster *Crassostrea gigas*. *Aquaculture International*, 25(4): 1507-1519. doi: 10.1007/s10499-017-0127-0
- Liu, Z. (2007). *Aquaculture genome technologies*. Blackwell Publishing Ltd, Oxford. doi: 10.1002/9780470277560
- Liu, Z.J. y Cordes, J.F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1-4): 1-37. doi: 10.1016/j.aquaculture.2004.05.027
- Lu, X., Kong, J., Meng, X., Cao, B., Luo, K., Dai, P. y 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 Immunology*, 73: 158-166. doi: 10.1016/j.fsi.2017.12.005
- Lyu, D., Yu, Y., Wang, Q., Luo, Z., Zhang, Q., Zhang, X., Xiang, J. y Li, F. (2021). Identification of growth-associated genes by Genome-Wide Association Study and their potential application in the breeding of Pacific white shrimp (*Litopenaeus vannamei*). *Frontiers in Genetics*, 12: 611570. doi: 10.3389/fgene.2021.611570
- Martinsohn, J.Th., Raymond, P., Knott, T., Glover, K.A., Nielsen, E.E., Eriksen, L.B., Ogden, R., Casey, J. y Guillen, J. (2019). DNA-analysis to monitor fisheries and aquaculture: Too costly? *Fish and Fisheries*, 20(2): 391-401. doi: 10.1111/faf.12343
- Massault, C., Jones, D.B., Zenger, K.R., Strugnell, J.M., Barnard, R. y Jerry, D.R. (2021). A SNP parentage assignment panel for the silver lipped pearl oyster (*Pinctada maxima*). *Aquaculture Reports*, 20: 100687. doi: 10.1016/j.aqrep.2021.100687
- Medrano-Mendoza, T., García, B.F., Caballero-Zamora, A., Yáñez, J.M., Montoya-Rodríguez, L., Quintana-Casares, J.C., Durán-Aguilar, M. y Campos-Montes, G.R. (2023). Genetic diversity, population structure, linkage disequilibrium and GWAS for resistance to WSSV in Pacific white shrimp (*Litopenaeus vannamei*) using a 50K SNP chip. *Aquaculture*, 562: 738835. doi: 10.1016/j.aquaculture.2022.738835
- Meehan, D., Xu, Z., Zuniga, G. y Alcivar-Warren, A. (2003). High frequency and large number of polymorphic microsatellites in cultured shrimp, *Penaeus (Litopenaeus) vannamei* [crustacea: decapoda]. *Marine Biotechnology*, 5(4): 311-330. doi: 10.1007/s10126-002-0092-z
- Mendoza-Cano, F., Grijalva-Chon, J.M., Pérez-Enríquez, R., Ramos-Paredes, J. y Varela-Romero, A. (2013). Genetic diversity of mitochondrial DNA from *Litopenaeus vannamei* broodstock used in northwestern Mexico. *Ciencias Marinas*, 39(4): 401-412
- Miller, L.M. y Kapuscinski, A.R. (2003). Genetic guidelines for hatchery supplementation programs. En: *Population genetics: Principles and applications for fisheries scientists*. American Fishery Society, Bethesda, pp. 329-355
- Nielsen, E.E., Cariani, A., Mac Aoidh, E., Maes, G.E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., Bekkevold, D., Diopere, E., Grenfell, L., Helyar, S., Limborg, M.T., Martinsohn, J.T., McEwing, R., Panitz, F., Patarnello, T., Tinti, F., Van Houdt, J.K.J., Volckaert, F.A.M., Waples, R.S., Albin, J.E.J., Vieites Baptista, J.M., Barmintsev, V., Bautista, J.M., Bendixen, C., Bergé, J.-P., Blohm, D., Cardazzo, B., Diez, A., Espiñeira, M., Geffen, A.J., Gonzalez, E., González-Lavín, N., Guarniero, I., Jérôme, M., Kochzius, M., Krey, G., Mouchel, O., Negrisolo, E., Piccinetti, C., Puyet, A., Rastorguev, S., Smith, J.P., Trentini, M., Verrez-Bagnis, V., Volkov, A., Zanzi, A. y Carvalho, G.R. (2012). Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, 3(1): 851. doi: 10.1038/ncomms1845
- Nolasco-Alzaga, H.R., Perez-Enriquez, R., Enez, F., Bestin, A., Palacios-Mechetnov, E. y 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. doi: 10.1016/j.aquaculture.2017.09.015
- Ogden, R. (2008). Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries*, 9(4): 462-472. doi: 10.1111/j.1467-2979.2008.00305.x
- Onda, Y., Takahagi, K., Shimizu, M., Inoue, K. y Mochida, K. (2018). Multiplex PCR targeted amplicon sequencing (MTA-Seq): Simple, flexible, and versatile SNP genotyping by highly

- multiplexed PCR amplicon sequencing. *Frontiers in Plant Science*, 9: 201. doi: 10.3389/fpls.2018.00201
- Palaiokostas, C. y Houston, R.D. (2017). Genome-wide approaches to understanding and improving complex traits in aquaculture species. *CABI Reviews*, 12(055). doi: 10.1079/PAVSNR201712055
- Peng, M., Chen, X., Yang, C., Liu, Q., Li, Q., Zhang, B., Wang, H., Zhu, W., Feng, P., Zeng, D. y Zhao, Y. (2023). A high-quality genome assembly of the Pacific white shrimp (*Litopenaeus vannamei*) provides insights into its evolution and adaptation. *Aquaculture Reports*, 33: 101859. doi: 10.1016/j.aqrep.2023.101859
- Peng, M., Zeng, D., Zhu, W., Chen, X., Yang, C., Liu, Q., Li, Q., Wang, H., Liu, H., Liang, J., Lin, Y., Chen, X. y Zhao, Y. (2020). Construction of a high-density genetic map and identification of Quantitative Trait Loci for nitrite tolerance in the Pacific white shrimp (*Litopenaeus vannamei*). *Frontiers in Genetics*, 11: 571880. doi: 10.3389/fgene.2020.571880
- Peñaloza, C., Manousaki, T., Franch, R., Tsakogiannis, A., Sonesson, A.K., Aslam, M.L., Allal, F., Bargelloni, L., Houston, R.D. y Tsigenopoulos, C.S. (2021). Development and testing of a combined species SNP array for the European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*). *Genomics*, 113(4): 2096-2107. doi: 10.1016/j.ygeno.2021.04.038
- Peñaloza, C., Robledo, D., Barría, A., Trinh, T.Q., Mahmuddin, M., Wiener, P., Benzie, J.A.H. y Houston, R.D. (2020). Development and validation of an open access SNP array for Nile Tilapia (*Oreochromis niloticus*). *G3: Genes Genomes Genetics*, 10(8): 2777-2785. doi: 10.1534/g3.120.401343
- Perez-Enriquez, R., Cruz-Hernández, P. y Saucedo-Barrón, C.J. (2013). Liberación de juveniles para la repoblación de cuerpos de agua, consideraciones y recomendaciones de manejo para minimizar los riesgos de impacto genético negativo. *Industria Acuícola*, 9(5): 18-23
- Perez-Enriquez, R., Hernández-Martínez, F. y 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
- Perez-Enriquez, R., Juárez, O.E., Galindo-Torres, P., Vargas-Aguilar, A.L. y Llera-Herrera, R. (2024a). Improved genome assembly of the whiteleg shrimp *Penaeus (Litopenaeus) vannamei* using long- and short-read sequences from public databases. *Journal of Heredity*, 115: 302-310. doi: <https://doi.org/10.1093/jhered/esae015>
- Perez-Enriquez, R., Llera-Herrera, R., Galindo Torres, P. y Avila, S. (2020). A fast sex detection method for the Whiteleg shrimp *Litopenaeus vannamei* by post-PCR high resolution melting (HRM). *Aquaculture*, 520: 734784. doi: 10.1016/j.aquaculture.2019.734784
- Perez-Enriquez, R., Magallón-Barajas, F.J. y Llera-Herrera, R. (2024b). Bases to inform a genetic line of whiteleg shrimp *Penaeus (Litopenaeus) vannamei* of Mexican origin. *Latin American Journal of Aquatic Research*, 52(1): 150-162. doi: 10.3856/vol52-issue1-fulltext-3066
- Perez-Enriquez, R. y Max-Aguilar, A. (2016). Pedigree traceability in whiteleg shrimp (*Litopenaeus vannamei*) using genetic markers: A comparison between microsatellites and SNPs. *Ciencias Marinas*, 42(4): 227-235. doi: 10.7773/cm.v42i4.2662
- Perez-Enriquez, R., Medina-Espinoza, J.A., Max-Aguilar, A. y Saucedo-Barrón, C.J. (2018a). Genetic tracing of farmed shrimp (Decapoda, Penaeidae) in wild populations from a main aquaculture region in Mexico. *Revista de Biología Tropical*, 66(1): 381-393. doi: 10.15517/rbt.v66i1.27112

- Perez-Enriquez, R., Millán-Márquez, A.M., Cruz-Hernández, P. y Saucedo-Barrón, C.J. (2018b). Population genetics of whiteleg shrimp *Litopenaeus vannamei* in Sinaloa, Mexico. *Revista Mexicana de Biodiversidad*, 89: 290-297. doi: 10.22201/ib.20078706e.2018.1.2070
- Perez-Enriquez, R., Robledo, D., Houston, R.D. y Llera-Herrera, R. (2018c). Genomics SNP markers for the genetic characterization of Mexican shrimp broodstocks. *Genomics*, 110: 423-429. doi: 10.1016/j.ygeno.2018.10.001
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. y Hoekstra, H.E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5): e37135. doi: 10.1371/journal.pone.0037135
- Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L. y Estoup, A. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, 95(6): 536-539. doi: 10.1093/jhered/esh074
- Prado, F., Vera, M., Hermida, M., Blanco, A., Bouza, C., Maes, G., Volckaert, F., Aquatrace, C. y Martínez, P. (2018). Tracing the genetic impact of farmed turbot *Scophthalmus maximus* on wild populations. *Aquaculture Environment Interactions*, 10: 447-463. doi: 10.3354/aei00282
- Premachandra, H.K.A., Nguyen, N.H. y Knibb, W. (2019). Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. *Aquaculture*, 499: 24-31. doi: 10.1016/j.aquaculture.2018.09.022
- Pritchard, J.K., Stephens, M. y Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2): 945-959
- Qi, H., Song, K., Li, C., Wang, W., Li, B., Li, L. y Zhang, G. (2017). Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). *PLoS ONE*, 12(3): e0174007. doi: 10.1371/journal.pone.0174007
- Rafalski, J.A. (2002). Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science*, 162(3): 329-333. doi: 10.1016/S0168-9452(01)00587-8
- Rannala, B. y Mountain, J.L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA*, 94(17): 9197-9201. doi: 10.1073/pnas.94.17.9197
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. y Mesirov, J.P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1): 24-26. doi: 10.1038/nbt.1754
- Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. y Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, 10(3): 670-682. doi: 10.1111/raq.12193
- Rodríguez-Ezpeleta, N., Díaz-Arce, N., Walter, J.F., Richardson, D.E., Rooker, J.R., Nøttestad, L., Hanke, A.R., Franks, J.S., Deguara, S., Lauretta, M.V., Addis, P., Varela, J.L., Fraile, I., Goñi, N., Abid, N., Alemany, F., Oray, I.K., Quattro, J.M., Sow, F.N., Itoh, T., Karakulak, F.S., Pascual-Alayón, P.J., Santos, M.N., Tsukahara, Y., Lutcavage, M., Fromentin, J.M. y Arrizabalaga, H. (2019). Determining natal origin for improved management of Atlantic bluefin tuna. *Frontiers in Ecology and the Environment*, 17(8): 439-444. doi: 10.1002/fee.2090
- Ruiz-Luna, A., Berlanga-Robles, C.A., Madrid Vera, J. y Rojas García, C.R. (2023). Pacific white shrimp (*Penaeus vannamei*) stock enhancement with aquaculture recruits: a case study in the coastal lagoons of southern Sinaloa, Mexico (2019-2020). *Latin American Journal of Aquatic Research*, 51(5): 671-684



- Santos, C.A., Rossini, B.C., Marques, C.G., Galetti, P.M. y Freitas, P.D. (2012). Characterization and genomic annotation of polymorphic EST-SSR loci in *Litopenaeus vannamei* shrimp. *Aquaculture Research*, 43(10): 1567-1570. doi: 10.1111/j.1365-2109.2011.02956.x
- Schmidt, D.A., Campbell, N.R., Govindarajulu, P., Larsen, K.W. y Russello, M.A. (2020). Genotyping-in-Thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Molecular Ecology Resources*, 20(1): 114-124. doi: 10.1111/1755-0998.13090
- Sellers, M.J., Dierens, L., McWilliam, S., Little, B., Murphy, B., Coman, G.J., Barendse, W. y Henshall, J. (2014). Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquaculture Research*, 45(3): 417-426. doi: 10.1111/j.1365-2109.2012.03243.x
- Shedd, K.R., Lescak, E.A., Habicht, C., Knudsen, E.E., Dann, T.H., Hoyt, H.A., Prince, D.J. y Templin, W.D. (2022). Reduced relative fitness in hatchery-origin Pink Salmon in two streams in Prince William Sound, Alaska. *Evolutionary Applications*, 15(3): 429-446. doi: 10.1111/eva.13356
- Soto-Rodriguez, S.A., Gomez-Gil, B., Lozano-Olvera, R., Betancourt-Lozano, M. y Morales-Covarrubias, M.S. (2015). Field and experimental evidence of *Vibrio parahaemolyticus* as the causative agent of acute hepatopancreatic disease of cultured shrimps (*Litopenaeus vannamei*) in northwestern Mexico. *Applied and Environmental Microbiology*, 81(5): 1689-1699. doi: 10.1128/AEM.03610-14
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., Jiang, C., Guan, N., Ma, C., Zeng, H., Xu, C., Song, J., Huang, L., Wang, C., Shi, J., Wang, R., Zheng, X., Lu, C., Wang, X. y Zheng, H. (2013). SLAF-seq: An efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE*, 8(3): e58700. doi: 10.1371/journal.pone.0058700
- Toonen, R.J., Puritz, J.B., Forsman, Z.H., Whitney, J.L., Fernandez-Silva, I., Andrews, K.R. y Bird, C.E. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1: e203. doi: 10.7717/peerj.203
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. y Rozen, S.G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40(15): e115. doi: 10.1093/nar/gks596
- Utter, F. (2003). Genetic impacts of fish introductions. En: *Population genetics: Principles and applications for fisheries scientists*. American Fisheries Society, Bethesda, pp 357-378
- Van der Pijl, W. (2020). *SPF L. vannamei Broodstock*. Shrimp insights, Netherlands. Disponible en: <https://shrimpsights.com/report-series/vannamei-broodstock> [consulta: 27 enero 2021]
- Vandeputte, M. y Haffray, P. (2014). Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. *Frontiers in Genetics*, 5: 432. doi: 10.3389/fgene.2014.00432
- Vandeputte, M., Rossignol, M.-N. y Pincent, C. (2011). From theory to practice: Empirical evaluation of the assignment power of marker sets for pedigree analysis in fish breeding. *Aquaculture*, 314(1-4): 80-86. doi: 10.1016/j.aquaculture.2011.01.043
- Vignal, A., Milan, D., SanCristobal, M. y Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3): 275-305. doi: 10.1051/gse:2002009
- Wang, Q., Yu, Y., Li, F., Zhang, X. y Xiang, J. (2017a). Predictive ability of genomic selection models for breeding value estimation on growth traits of Pacific white shrimp *Litopenaeus vannamei*. *Chinese Journal of Oceanology and Limnology*, 35(5): 1221-1229. doi: 10.1007/s00343-017-6038-0

- Wang, Q., Yu, Y., Yuan, J., Zhang, X., Huang, H., Li, F. y Xiang, J. (2017b). Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. *BMC Genetics*, 18: 45. doi: 10.1186/s12863-017-0507-5
- Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Huang, H., Xiang, J. y Li, F. (2019a). Evaluation on the genomic selection in *Litopenaeus vannamei* for the resistance against *Vibrio parahaemolyticus*. *Aquaculture*, 505: 212-216. doi: 10.1016/j.aquaculture.2019.02.055
- Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Yuan, J., Huang, H., Xiang, J. y Li, F. (2019b). A novel candidate gene associated with body weight in the pacific white shrimp *Litopenaeus vannamei*. *Frontiers in Genetics*, 10: 520. doi: 10.3389/fgene.2019.00520
- Wang, S., Meyer, E., McKay, J.K. y Matz, M.V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, 9(8): 808-810. doi: 10.1038/nmeth.2023
- Weir, L.K. y Grant, J.W. (2005). Effects of aquaculture on wild fish populations: a synthesis of data. *Environmental Reviews*, 13(4): 145-168. doi: 10.1139/a05-012
- Wolfus, G.M., Garcia, D.K. y Alcivar-Warren, A. (1997). Application of the microsatellite technique for analyzing genetic diversity in shrimp breeding programs. *Aquaculture*, 152(1-4): 35-47. doi: 10.1016/S0044-8486(96)01527-X
- Yáñez, J.M., Naswa, S., López, M.E., Bassini, L., Correa, K., Gilbey, J., Bernatchez, L., Norris, A., Neira, R., Lhorente, J.P., Schnable, P.S., Newman, S., Mileham, A., Deeb, N., Di Genova, A. y Maass, A. (2016). Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular Ecology Resources*, 16(4): 1002-1011. doi: 10.1111/1755-0998.12503
- Yu, Y., Wang, Q., Zhang, Q., Luo, Z., Wang, Y., Zhang, X., Huang, H., Xiang, J. y Li, F. (2019). Genome scan for genomic regions and genes associated with growth trait in Pacific white shrimp *Litopenaeus vannamei*. *Marine Biotechnology*, 21: 374-383. doi: 10.1007/s10126-019-09887-w
- Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., Huang, L., Zheng, H. y 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
- Zenger, K.R., Khatkar, M.S., Jones, D.B., Khalilisamani, N., Jerry, D.R. y Raadsma, H.W. (2019). Genomic Selection in aquaculture: Application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Frontiers in Genetics*, 9: 693. doi: 10.3389/fgene.2018.00693
- Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., Liu, C., Wang, Q., Lv, X., Zhang, X., Ma, K.Y., Wang, X., Lin, W., Wang, L., 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, L., Ruan, J., Chu, K.H., Liu, B., Li, F. y Xiang, J. (2019). Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. *Nature Communications*, 10(1): 356. doi: 10.1038/s41467-018-08197-4
- Zhang, Z., Glover, K.A., Wennevik, V., Svåsand, T., Sørvik, A.G.E., Fiske, P., Karlsson, S. y Skaala, Ø. (2013). Genetic analysis of Atlantic salmon captured in a netting station reveals multiple escapement events from commercial fish farms. *Fisheries Management and Ecology*, 20: 42-51. doi: 10.1111/fme.12002
- Zhao, H., Li, C., Hargrove, J.S., Bowen, B.R., Thongda, W., Zhang, D., Mohammed, H., Beck, B.H., Austin, J.D. y Peatman, E. (2018). SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*). *Aquaculture*, 485: 30-38. doi: 10.1016/j.aquaculture.2017.11.014

Zhi-Ying, J., Xiao-Wen, S., Li-Qun, L., Da-Yu, L. y Qing-Quan, L. (2006). Isolation and characterization of microsatellite markers from Pacific white shrimp (*Litopenaeus vannamei*). *Molecular Ecology Notes*, 6(4): 1282-1284. doi: 10.1111/j.1471-8286.2006.01515.x

## 11. ANEXOS

### Anexo A. Comandos utilizados para identificar los sitios de reconocimiento de enzimas IIB

Los complementos *-A* y *-B* del comando *zgrep* indican que se obtienen los renglones después y antes del renglón de interés, respectivamente. Los comandos *paste* y *sed* sirven para eliminar espacios, retornos de carro y guiones para que cada sitio reconocido quede en una sola línea. El comando *cat* al final une a cada uno de los archivos obtenidos en uno solo.

# Código para la búsqueda y conteo de sitios de restricción de la enzima *Bcgl* con reconocimiento de la secuencia CGA.....TGC en el genoma de *P. vannamei* (GCA\_003789085.1\_ASM378908v1\_genomic.fna.gz) donde los puntos indican cualquier base. Con el fin de contar los sitios que quedan al final y al principio de los renglones se incluyen las diversas posibilidades usando los símbolos *\$* al final del renglón y *^* para el principio del renglón.

```
zgrep -A 1 -B 1 'CGA.....TGC' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' > sitios_Bcgl_1.txt
```

```
zgrep -A 1 'CGA.....TG$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_2.txt
```

```
zgrep -A 1 'CGA.....T$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_3.txt
```

```
zgrep -A 1 'CGA.....$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_4.txt
```

```
zgrep -A 1 'CGA.....$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_5.txt
```

```
zgrep -A 1 'CGA....$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_6.txt
```

```
zgrep -A 1 'CGA...$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_7.txt
```

```
zgrep -A 1 'CGA..$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_8.txt
```

```
zgrep -A 1 'CGA.$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_9.txt
```

```

zgrep -A 1 'CGA$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' |
sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_10.txt
zgrep -B 1 '^A.....TGC' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed
's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_11.txt
zgrep -B 1 '^GA.....TGC' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed
's/\t//g' | sed 's/--\n/g' | grep 'CGA.....TGC' > sitios_Bcgl_12.txt
cat sitios_Bcgl_1.txt sitios_Bcgl_2.txt sitios_Bcgl_3.txt sitios_Bcgl_4.txt sitios_Bcgl_5.txt
sitios_Bcgl_6.txt sitios_Bcgl_7.txt sitios_Bcgl_8.txt sitios_Bcgl_9.txt sitios_Bcgl_10.txt
sitios_Bcgl_11.txt sitios_Bcgl_12.txt > sitios_Bcgl_Fwdtotal.txt
# Esto también debe ser calculado para la secuencia antisentido que es 'GCT.....ACG', pero dado
#que la secuencia solo la tenemos en dirección sentido, se debe buscar en dirección inversa es
#decir 'GCA.....TCG'
zgrep -A 1 -B 1 'GCA.....TCG' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed
's/\t//g' | sed 's/--\n/g' > sitios_Bcgl_Rev1.txt
zgrep -A 1 'GCA.....TC$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed
's/\t//g' | sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev2.txt
zgrep -A 1 'GCA.....T$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed
's/\t//g' | sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev3.txt
zgrep -A 1 'GCA.....$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g'
| sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev4.txt
zgrep -A 1 'GCA.....$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g'
| sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev5.txt
zgrep -A 1 'GCA....$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g'
| sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev6.txt
zgrep -A 1 'GCA...$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' |
sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev7.txt
zgrep -A 1 'GCA..$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' |
sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev8.txt
zgrep -A 1 'GCA.$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' |
sed 's/--\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev9.txt

```

```
zgrep -A 1 'GCA$' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed 's/\t//g' |  
sed 's/--/\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev10.txt  
zgrep -B 1 '^A.....TCG' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed  
's/\t//g' | sed 's/--/\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev11.txt  
zgrep -B 1 '^CA.....TCG' GCA_003789085.1_ASM378908v1_genomic.fna.gz | paste -s | sed  
's/\t//g' | sed 's/--/\n/g' | grep 'GCA.....TCG' > sitios_Bcgl_Rev12.txt  
cat sitios_Bcgl_Rev1.txt sitios_Bcgl_Rev2.txt sitios_Bcgl_Rev3.txt sitios_Bcgl_Rev4.txt  
sitios_Bcgl_Rev5.txt sitios_Bcgl_Rev6.txt sitios_Bcgl_Rev7.txt sitios_Bcgl_Rev8.txt  
sitios_Bcgl_Rev9.txt sitios_Bcgl_Rev10.txt sitios_Bcgl_Rev11.txt sitios_Bcgl_Rev12.txt >  
sitios_Bcgl_Revtotal.txt
```

## Anexo B. Secuencias de adaptadores y oligos para preparación de librerías 2bRAD

Secuencias de adaptador 1 y adaptadores 2 (2.1 al 2.12, 12 posibles adaptadores para cada columna de la placa) en sentido y su correspondiente antisentido, sin bases selectivas. En rojo, bases que pueden ser substituidas con bases selectivas; en azul, secuencias de código de barras.

ID Adaptador	Secuencia
Adaptador1_5ill-NNRW	CTACACGACGCTCTTCCGATCTNNRWCCNN
Adaptador1_anti5ill-NNRW	GGWYNNAGATCGG/3InvdT/
Adaptador2.1_3illBC1	CAGACGTGTGCTCTTCCGATCTACACNN
Adaptador2.1_antiBC1	GTGTAGATCGGA/3InvdT/
Adaptador2.2_3illBC2	CAGACGTGTGCTCTTCCGATCTGTCTNN
Adaptador2.2_antiBC2	AGACAGATCGGA/3InvdT/
Adaptador2.3_3illBC3	CAGACGTGTGCTCTTCCGATCTTGGTNN
Adaptador2.3_antiBC3	ACCAAGATCGGA/3InvdT/
Adaptador2.4_3illBC4	CAGACGTGTGCTCTTCCGATCTCACTNN
Adaptador2.4_antiBC4	AGTGAGATCGGA/3InvdT/
Adaptador2.5_3illBC5	CAGACGTGTGCTCTTCCGATCTGATGNN
Adaptador2.5_antiBC5	CATCAGATCGGA/3InvdT/
Adaptador2.6_3illBC6	CAGACGTGTGCTCTTCCGATCTTCACNN
Adaptador2.6_antiBC6	GTGAAGATCGGA/3InvdT/
Adaptador2.7_3illBC7	CAGACGTGTGCTCTTCCGATCTCTGANN
Adaptador2.7_antiBC7	TCAGAGATCGGA/3InvdT/
Adaptador2.8_3illBC8	CAGACGTGTGCTCTTCCGATCTAAGCNN
Adaptador2.8_antiBC8	GCTTAGATCGGA/3InvdT/
Adaptador2.9_3illBC9	CAGACGTGTGCTCTTCCGATCTGTAGNN
Adaptador2.9_antiBC9	CTACAGATCGGA/3InvdT/
Adaptador2.10_3illBC10	CAGACGTGTGCTCTTCCGATCTGACANN
Adaptador2.10_antiBC10	TGTCAGATCGGA/3InvdT/
Adaptador2.11_3illBC11	CAGACGTGTGCTCTTCCGATCTGTGANN
Adaptador2.11_antiBC11	TCACAGATCGGA/3InvdT/
Adaptador2.12_3illBC12	CAGACGTGTGCTCTTCCGATCTAGTCNN
Adaptador2.12_antiBC12	GACTIONAGATCGGA/3InvdT/

Secuencias de oligos ILLBC (ILLBC23-30, ocho posibles oligos para cada línea de la placa); Truseq (TruseqUn1-6, seis posibles oligos para identificar cada placa); P5 y P7 (oligos para secuenciación con Illumina). En rojo, la secuencia del código de barras.

ID oligo	Secuencia
ILLBC23	CAAGCAGAAGACGGCATAACGAGAT <b>CCACTC</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC24	CAAGCAGAAGACGGCATAACGAGAT <b>GCTACC</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC25	CAAGCAGAAGACGGCATAACGAGAT <b>ATCAGT</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC26	CAAGCAGAAGACGGCATAACGAGAT <b>GCTCAT</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC27	CAAGCAGAAGACGGCATAACGAGAT <b>AGGAAT</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC28	CAAGCAGAAGACGGCATAACGAGAT <b>CTTTTG</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC29	CAAGCAGAAGACGGCATAACGAGAT <b>TAGTTG</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
ILLBC30	CAAGCAGAAGACGGCATAACGAGAT <b>CCGGTG</b> GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
Truseq_Un1	AATGATACGGCGACCACCGAGATCTACAC <b>ATCACG</b> AACACTCTTTCCCTACACGACGCTCTTCCGATCT
Truseq_Un2	AATGATACGGCGACCACCGAGATCTACAC <b>ACTTGA</b> AACACTCTTTCCCTACACGACGCTCTTCCGATCT
Truseq_Un3	AATGATACGGCGACCACCGAGATCTACAC <b>TAGCTT</b> AACACTCTTTCCCTACACGACGCTCTTCCGATCT
Truseq_Un4	AATGATACGGCGACCACCGAGATCTACAC <b>GGCTAC</b> AACACTCTTTCCCTACACGACGCTCTTCCGATCT
Truseq_Un5	AATGATACGGCGACCACCGAGATCTACAC <b>TAGGCA</b> AACACTCTTTCCCTACACGACGCTCTTCCGATCT
Truseq_Un6	AATGATACGGCGACCACCGAGATCTACAC <b>CGATGT</b> AACACTCTTTCCCTACACGACGCTCTTCCGATCT
P5	AATGATACGGCGACCACCGA
P7	CAAGCAGAAGACGGCATAACGA



### **Anexo C. Descripción de análisis de datos generados durante la qPCR-HRM para curvas de alta resolución (HRM) y obtención de genotipos de SNPs**

El software del termociclador CFX96 genera los datos de la fluorescencia en RFU (Unidades de Fluorescencia Relativa) en función de la temperatura durante el análisis de la curva de disociación en un archivo que nombra "*Melt curve amplification results*", además de la temperatura inicial y final de cada curva de disociación en el archivo "*Melt curve peak results*" y la derivada de la fluorescencia en función de la temperatura en el archivo "*Melt curve derivative results*". Los datos se exportan a Excel donde inicialmente son graficados considerando el rango de temperatura de la curva dado por el equipo. Se acota el rango de temperatura del punto previo al descenso la fluorescencia hasta que desciende completamente. Este rango representa del punto previo al inicio de la disociación de los fragmentos de ADN hasta el punto en que todos se encuentran separados. Este intervalo de datos es usado en el cálculo de las curvas subsecuentes.

Se calculan tres curvas para cada SNP: (1) Una curva normalizada para incrementar inicialmente la diferencia entre los perfiles de las curvas, en esta los datos son normalizados en valores relativos de 1 a 0. (2) Una curva diferenciada con el objetivo de enfatizar la diferencia en el perfil de cada uno de los tres genotipos, en la cual (después de la normalización) se promedian los datos de un cluster seleccionado, definiendo como cluster las curvas que están agrupadas por ser muy similares y que presuntamente corresponden a un genotipo; los datos de todos los individuos son diferenciados del promedio del cluster. (3) Y una curva derivativa normalizada que representa la variación de la fluorescencia, en la cual la derivada de la fluorescencia en función de la temperatura es graficada y luego normalizada en valores de 1 a 0, esta curva en su pico más alta representa el punto teórico en el que el 50% de los fragmentos de ADN están disociados.

#### Anexo D. Temperatura de alineamiento y secuencia de oligos para HRM del panel de SNPs para análisis de estructura genética y asignación poblacional

Secuencia de los oligos diseñados para cada SNP y la temperatura final de alineamiento utilizada durante la qPCR. Se incluyen sólo los 35 marcadores moleculares que conforman el panel que fue utilizado en los análisis *in silico*.

Panel de 35 SNPs. (F) secuencia sentido; (R) secuencia antisentido; (TA) Temperatura de alineamiento.

ID SNP	Secuencia oligos	TA °C
Lv24611_43_128	F: GGTTTAAGGTAAGCCCCGTA R: TGGTATTTTTGCAGGTATTG	55
Lv5199_109_121	F: CATTGGGAATGCATGGCGG R: TTTCATTCTCCCTCCACGCG	57
Lv11392_361_39	F: ACCAGGAATCACTTTGCTTGC R: TGTGCATGCAACTGCTTCAG	-
Lv5830_30_82	F: TCTTTGGGAAAGGAGTCGCC R: TGGGAGAACTGGATTAAAGTATCC	57
Lv3960_43_46	F: TTCTCTCCACATCCTGTGCGC R: ACCAACAAGAGCCATTTACCCT	57
Lv26562_49_27	F: TCAGAGGAACCGGGTAGTGA R: CTCACTCGCCGCACTTCG	57
Lv2_10168_152_75	F: ACCCAAGCTGCCTCAAAAGT R: AAATACAGGGGGCAAGTACG	55
Lv2_7839_114_72	F: GAGGAACCTCGCGTCGATG R: GTGACGTCACTGACCTGCTG	57
Lv2_8274_40_59	F: CACCATGAACACGTTCTGCA R: TCGTGGATCAACTGCGTGAA	55
Lv2_20522_88_45	F: AGCCTCGTTAGCAGACGTTT R: AGCCTTTCTAACATCGTCTAACA	57
Lv2_759_312_40	F: AACCCACGTCACCATGAACA R: ACTTGAGCGAAACATTACG	55
Lv2_2539_112_25	F: TGATTCAGTTGCCGAACCCA R: GGAGATCGCACCGCTAACTT	57

Lv2_3213_36_42	F: GGAGTCAGTGTGTGGGGAAT R: AGGTACACAGGGAGAAACCT	57
Lv2_2358_214_66	F: TTA ACTCGTCAGCCGGCTTT R: GAGGGGTCAAGCCTCGAATG	57
Lv2_2355_1228_73	F: TGACCTGTTTAGTGCCCTGG R: TCACCACCGGCCAGATATTG	57
Lv2_2849_118_101	F: ACCATGACTTTTCATCTTCCATCT R: TGTGCTCTCTCAGGACCTCT	57
Lv3500_48_64	F: CTCGCGTGTAACACCGGAAG R: ACTGTAGACATACACGTGCACA	59
Lv7641_72_71	F: CGTCCACCACCTCAACATCA R: GGCGATGATGACCTTGCCA	57
Lv9003_194_99	F: AGCCAACACCTTAAACAGCA R: GCGGTCTTCGTACGAGTTCA	55
Lv22879_167_52	F: CGGCTCGCTACGAAATCACT R: GATGAATTGTCGCGCCTTG	57
Lv26271_31_118	F: CGTTGCTTGAGGAACTGACG R: GTCCTCCGGCTAAAAGTCCC	59
Lv513_155_122	F: ATAGGCCCTGTCTTGCGTTT R: TCCTTCATAGATCGTTTCCCACA	59
Lv9814_36_52	F: ATGTCTTTGTGGTTTGTAGAAAAT R: TAGATTTACAGTCTATCGCCTACACC	59
Lv8647_40_42	F: CAGTCAGGGCCTCAAGCTC R: TCCTGTAGCCGCCTTCTTTC	57
Lv25396_159_54	F: ATCTGTGTGCAAAGACATTTAACA R: ATGCCATCCAACATCTTCTACCT	59
Lv10866_351_29	F: CTTATGGACGGTTGGGCGG R: TCTCGCTCCCTCTCTCTGTT	57
Lv8642_60_52	F: ACGCAGAACTAGACTTTGGGT R: AATCGAATGGCGGAGTAGG	57
Lv7158_209_86	F: AACAAACACACTAGACCGGGC R: GTAAAGGTACTGGCAGGGTCC	59

Lv5626_1772_46	F: GTGTAGACCCGTGGCTCAG R: TTCTTTCGCACCTCTGTCCC	57
Lv5234_134_32	F: GGGGTCGGCAATGTGTACAA R: CCCCCTTTTGCTTTATATCATGT	59
Lv4410_124_31	F: AGGGTACTGATATGGATAAGGA R: CTCCTGTCACGCGAAGAAGT	59
Lv4048_586_118	F: GAGGCCACATCATAACCGCT R: CGGGTTTTCTGGCTTAGCT	57
Lv3322_409_22	F: GTATACCACAAAGGGAGCAAG R: GGTTGGTTTCTTTTCTCCGC	57
Lv2700_99_106 <sup>1</sup>	F: CCTTCAGGTTGTACGGCCC R: AGTTTTCAGTTCCTATAATTCACATGA	-
Lv1277_233_73	F: TGTTTGATTTGGGCCGACCT R: AGGCACATCACTTCCTTCGG	57

---

<sup>1</sup>No amplificó en qPCR

### Anexo E. Panel de 118 SNPs para estructura genética y asignación poblacional, análisis de parentesco e identificación de sexo.

Los oligos diseñados para este panel pueden ser incluidos en una sola reacción multiplex para la obtención de genotipos por GT-seq. Cabe recordar que esta es la secuencia locus-específico, al oligo es necesario adicionarle las secuencias necesarias para su genotipificación por GT-seq, tal como se menciona en la sección 6.1.2.2 y 6.3.2 de Materiales y Métodos.

Panel de 118 SNPs para análisis de estructura genética y asignación poblacional, análisis de parentesco e identificación de sexo. (F) secuencia sentido; (R) secuencia antisentido; (GP) SNP seleccionado para estructura genética y asignación poblacional; (P) SNP seleccionado para análisis de parentesco; (IDS) SNP para identificación de sexo.

ID SNP	Secuencia oligos	Tipo de SNP
26659_141_123	F: TGAGGGGTTGCGCAATACTT R: AACTCCGTCTCGCGTAGA	P
26358_32_62	F: CAACAGAGGGCCCGAATCG R: GAATGAATCTCCCGCCGAA	P
25452_39_86	F: CTTGAGACGCTCTCGCTAGG R: GCTGCGAGTTGAAGTCTCCA	P
25084_128_106	F: ATGCCGGGCCAGAATAATGT R: GTGCGGCTATATTTACACGAGC	P
23987_42_95	F: GCTTCTCCTTGTCTGTGCT R: ACAACAAGACTGAGCAGGGT	P
23612_43_30	F: CGTGATAAAAGAACTTTCTATTTTGCG R: CGAAAGCAAGAACCGCTACG	P
23241_32_89	F: TCCATCCTCTCGTTCGCTTG R: AAAGCCGACAGACATTCCCA	P
23179_194_57	F: GCCTCTCAAGCTGACAGTACA R: TGACGAACAGTGTTCCAATTGA	P
23048_265_32	F: TGGTAGGATCTAAAAGTGTCACAC R: CACTTTCGCCAGTCACAACG	P
22476_60_84	F: GCCAAGAAGGTGTTGGGAGA R: GGCCGCTTATCTTTTTCCCC	P
22254_93_66	F: CATCAGCAAAGTCAAGGTCAGT	P

	R: GCAACAACACCTTTGCATGGA	
21988_37_80	F: TGAACTCACCCAATACCCCC R: GTCACAGGGAAGGAAAAGGGA	P
21571_174_67	F: TGTTGACGGTCCTTCCTGTG R: GTAGCTTTGGGAACGGTGGA	P
21052_66_37	F: TGCCCCAGTGTTACTCCCTA R: TTCCTGTTCCCCAGACGCTA	P
20566_45_29	F: AGAGTATCCTGCGACTCCCC R: TCCTAGCTGGGACGAGGATG	P
19765_60_72	F: GGAATGCTGTCACGTGAAAGG R: CCAAGGGCGTTCGAAACTTG	P
19680_80_76	F: ACCTTCTCTCGGCACCATTG R: TCCTTTTTACGCTCCGCCAT	P
19529_54_78	F: TAGGAGATGCTAGGGCCGAG R: ACAAGCACAGCACTTCCTCA	P
19379_31_100	F: TCTTGCCTCTCTGAGCAGC R: GACTCTCAGTCAATTCAATCTCGA	P
19292_52_31	F: TCTGTGCACTGCCCTTCG R: AATGGCACTGCGGTGAGAAG	P
19183_297_42	F: GGACAGCTCCCTGGTCAAAG R: GGCTCTTTGGCAGGTTTTGG	P
18433_37_100	F: TCAGAGTCCCCATCCCCTT R: TGAGGCCTTAGACTGTTGCG	P
17910_36_120	F: CGCTATTGTTGGGGGTCCT R: CCAGCCACACCTCCCCTC	P
17883_117_37	F: CGTGAGGGTTGGGTTCTCTC R: CACCCAGTGTGCCTGTA ACT	P
17752_60_83	F: CCATACAGCACGGGGTTGAT R: TTACCGTGGCATGTCG TCAA	P
17577_54_100	F: AGAGAGAGGGAGTGCAGGAG R: CCTCCCGTTCTTCTCCTTCG	P
16885_30_29	F: GTACCCAGGTGAGCCGC	P

	R: CGCAAGGGTTGTTTGGCTTT	
16778_168_114	F: TGTCTCCCGCGAGTCTCATA R: TTGCCATGACGCTGACCTG	P
16648_417_58	F: TCACGGCATCTGTGAAGTCA R: AGCGAGGCAGCAAACCTACT	P
16534_42_89	F: CTCCGGCAACCATCCAATCT R: TTTCACGTCAATCTTCCATCCAT	P
16243_36_86	F: TCAGCAAAGTCAAGGTCGGT R: ACACTTGCATGGACTGGACG	P
15250_39_38	F: ATCCGGAGCAGAGTTTACCC R: AGGAGGAAAGCGAGTCATGC	P
15073_47_111	F: CAGGTTTTGCCTCCTCCTCC R: GCCACTTGTTGCTTCTGC	P
14885_52_87	F: CCCTGTGTCTGGAGGTTTGT R: CCTTAAAGCTCACGTGGCAC	P
14681_61_74	F: ACATTTCACTTGTGTACCAGGGA R: ATTGGCTGATGGTCAGGGTG	P
14623_62_83	F: AACGACAAGACGCGGTTTTG R: TTAGGTCCGACTCGCATGAC	P
14538_61_117	F: CAGGTCATAGCCCTTCCACG R: GTCGTCCGAAGTGTGCCG	P
14108_1170_97	F: CAAAGAAGGTTGGGGCAAGC R: ATGTGGAGAGTGGCGTGAAG	P
14032_58_121	F: TGAGCCAATGAGTCAGCGAC R: TCGTTAAAGCCTTCCCTCGC	P
13056_69_23	F: ACAGGCGGTAAGTATGCAGC R: TGCTGGATAATAGGACGGGGA	P
13007_59_30	F: TCGTCTTCTGTTGTACATCGTTT R: TTGAACACTTCCCTCGTCGCC	P
12523_426_47	F: CCCCCAATCCCACTTCTCTG R: GTGACAGCACTCTCGCAGAT	P
12314_124_128	F: TTTAGTGTGTGACGCCCTCG	P

	R: GAAGCGGTGATCTGAGGAGC	
11321_70_27	F: CTCCCAGATTTGAGGAGCCG R: ACATACATCACTGCTGCTGTCA	P
11208_39_117 <sup>1</sup>	F: CAGATGCCATTCCAGGACGA R: ACCCTGGCCGGTTTATCAAG	P
10925_305_127	F: CACCGTAAGAGCATGTCCA R: TGAGATTCCTGCTGGTGAGG	P
10805_123_53	F: ATGCCCGAAGTTTGGAGGAA R: TTGTCTGCTATTGTAAGTGGCTT	P
9717_50_71	F: ACGGAGTGAAATAGTACATTTTGAATG R: CGTCGCCACCGTTATACAT	P
7606_31_72	F: GCACTTGCGATGGTAAAATCATTG R: AGCTCCACAGTGA CTTCCTT	P
7515_119_39	F: TGTA ACTGTGCAGTGGACGC R: GGATTCGTCTCCAGCCTCAG	P
7257_133_89	F: GGATCCTTTTTAAAACCCAGTGCT R: AGCAGGGTTTTGCCTAAACAA	P
7042_73_65	F: ATATGTCCGCGGGCTTTGAT R: GCCAACC ACTAACCACACT	P
6795_169_62	F: GCGTGGA AATGGAGGGGAAA R: TGCATCTCTCTCGCCCTTTG	P
6602_50_98	F: GAGTCTTATCCGGGTCGCAG R: ACACCTCAGCCTTCAACACC	P
6483_44_60	F: CAGTCTCAGAGTCGCCAAGG R: GATCACCTGGCTTCGTTTGC	P
6189_63_60	F: CAAGATCGCCGAGTTCTCGA R: AA ACTGACCTGAACTCGGGC	P
5861_232_119	F: AAGTTCTCGAAAGAAGGCGG R: GCTCCGTCGTCTCGTAATGT	P
5777_55_112	F: TGCCCTTGTTGGAGAGCATT R: TGTGGCAATAATGAAAATAAGTGACA	P
5574_49_23	F: GCGAGATGGACCGAGGTTTC	P



	R: CCCTTCCTCTGCCCTCCATA	
5477_81_39	F: GTGATCACACTGGTACATGATATTTGT R: GCCACCTGCACCATCATACT	P
5234_134_32	F: GGGGTCGGCAATGTGTACAA R: TCCCCCTTTTGCTTTATATCATGT	P
5197_37_101	F: CGTCCTCCTTGTGGGTGAAA R: TCCTTTCTTTCTCTCCTTGTCTCG	P
5123_49_69	F: GACCTCCAGCAGCACAACAA R: GTCTGGGGCGATATGCTCTC	P
2886_253_38	F: CAGGCTGTATGGCAGGGTCGC R: TCACCCCAAGCCACGCTCT	P
4335_52_100	F: TCTGGCACTCTGCGGGAGGC R: AGCTCATGGGCGGCTGGTAG	P
4084_41_118	F: CAACCCGTTTGCCGCAACG R: AGGTGTGGGGTGGGTGCGAGG	P
3494_140_31	F: GGGACACCTGAGCCACATC R: TTTCTCTCGAGGGAGCGTTG	P
2553_114_43	F: CCTCCATCACCTCATCAGC R: GATACAGCGCACGACCTACA	P
2539_112_25	F: TGATTCAGTTGCCGAACCCA R: AGCTGGGATGGCGCATATAC	P
2449_63_109	F: TGTCTGGTTTTGCCCACT R: GACTGGCGTAAGTTTGCACA	P
2346_202_39	F: AGACGTTTCAGATGGCTCGAG R: GACAACGTTTCAGCCATGCAA	P
2145_32_41	F: AATCAGCAGGGCGCCGAG R: TCACATCCTCTGGGTTGCTG	P
17599_44_83	F: TGTCGGAGGGGTTATGAGGT R: ATGCACTCGTCAGCAGGC	P
7087_47_70	F: GACCTCCACCCCTACACCC R: CTCTCCTGCCGGACGTCC	P
5500_31_87	F: CGAAGGTGAGAGTGAGAGCG	P

	R: TGACGTCGGCCTCCTGTC	
9223_34_81	F: AGTTATCTGCGTGGGACAGC R: CGCTCGTTCGCTCATGCTAT	P
6175_66_115	F: TTCGCAGTTTAAAATATTGGAAGCA R: TTTTGTGAGGTGATTTTCTTCCTT	P
287_30_81	F: CGACTCCCTCCCTCTTTGA R: TTA CTCAAGCAGGACAGCCG	P
8953_32_69	F: TCGGTTGCATATATCATTCCCA R: AGACAATTCATAGACCGAGAAAGAGA	P
12458_182_40	F: TACCACCTCGAGCCTCTGG R: CCCGAGATGAATCCCATGGG	P
1943_140_84	F: CCGCGATTTATTTGGCGTGA R: ACGCCAAATGATGTTTCGCC	P
3467_56_72	F: TGGTCTTCCTGGGTGAGTCA R: ACGTCTCCAGGAACCGAAAC	P
207_101_94	F: TGGGCATAAAGGTTCTTATCAACC R: CCAGTTCTCAAGTATGATCGATGG	P
11550_33_63	F: CTTTCGCTGGGAGGACCTTT R: CGGGCTCAGGAGTAAGCAA	P
20437_30_59	F: GAACAGCTTCTGGGAGCCAT R: GTACCCAGCCATCTGACAGA	P
10471_66_124 <sup>3</sup>	F: GGGTGGTGAGTCAAAGTGCT R: CCAGAACTATGGCCCAGTCA	P
2933_38_119	F: TCTGAACACCAACTTGGGCA R: CCTCCCAATAGGTGTGCAA	P
12227_85_38	F: CACGCACATGCACATCCAC R: CGCATAAAAGGCAGCGCATT	P
23746_40_91	F: ATTTCCATGACTCACGCCGT R: CACCGGAGAGACCCTGTTTC	P
5309_122_48	F: GCTGATCGGAAGCTTCTCGA R: GACTTACTCGGAGCGCAGG	P
4424_107_115	F: GGTTGCTTCAAGTGCCATT	P

	R: GTGATTGGCACGGATTCCC	
4532_52_56	F: CAGCCTCGCCTCTCACATAA R: TCCCGCATTACACAACACAAC	P
32_179_24	F: CGCCGCACGTAATAAAGCA R: TGATGGCGGTTTGTCTTGGA	P
953_97_86	F: AGGCGATGAGGAGGGAGTAG R: TCCTCTGCCATTATTTCAGCACT	P
18219_42_111	F: CGACGTCGTTTGTGAGGGAT R: ATTAGCACCCCTCCCGAGAA	P
4396_49_88	F: GTGCCCTGTTTCTCTCCCTC R: AGAGGTGACAACGCTTCTAAGA	P
5251_41_29	F: TGAAAGCTTCTTCGGAAACACG R: GCAGCTAAATCAACCACCGC	P
2358_214_66	F: GAGCGCATTAACTCGTCAGC R: GCTCTGGAGACGATTTCCCG	P
20581_53_46	F: CACGGGTCGGACTCGAAC R: ACGATGGCCTTCGCTGTC	P
14061_31_76	F: TGAGGATGAGCACCTTGAGC R: CTGTCGCTCTCGTGGCTG	P
26271_31_118	F: GCTTGAGGAACTGACGGGAA R: GTCCTCCGGCTAAAAGTCCC	GP
759_312_40	F: AACCCACGTCACCATGAACA R: GCAACAGAACACACGGTGGA	GP
25396_159_54	F: TCTGTGTGCAAAGACATTTAACAAA R: AGAACGGAAAGTGAGTGGACT	GP
22879_167_52	F: GAGAACACCTAACCGCAGCT R: AATAGCGCTGCTTCCCTTGT	GP
3213_36_42	F: GGAGTCAGTGTGTGGGGAAT R: ACTTTGGTCATGTCAGCCCT	GP
8274_40_59	F: GGCCACGATGATCATGACCA R: TCGTGGATCAACTGCGTGAA	GP
3960_43_46	F: TTCTCTCCACATCCTGTCGC	GP

	R: CGAGGCCGAGTCTTTACAGG	
2849_118_101	F: TCAAGGATCCCATTACCATGACT R: TGTGCTCTCTCAGGACCTCT	GP
20522_88_45	F: AGCCTCGTTAGCAGACGTTT R: CCCGCCTTCATATCAGCTTGA	GP
5830_30_82	F: TCTTTGGGAAAGGAGTCGCC R: GTGATGTTGTGGGCGTTTC	GP
10168_152_75	F: ACCCAAGCTGCCTCAAAAGT R: CCCGTTGTCTTTCTCCTTCGA	GP
26562_49_27	F: TCAGAGGAACCGGGTAGTGA R: GACGAGCAGCTCCTTCCTTC	GP
11392_361_39	F: ACCAGGAATCACTTTGCTTGC R: TGTGCATGCAACTGCTTCAG	GP
5199_109_121	F: CAGGTCACCCTCCTTCACAC R: TTTCATTCTCCCTCCACGCG	GP
9814_36_52 <sup>1</sup>	F: GGTTTGTAGAAAATAATCATCGTTGCG R: CTATCGCCTACACCTCAGCC	GP
513_155_122	F: ATAGGCCCTGTCTTGCGTTT R: TCTCCTTCATAGATCGTTTCCCAC	GP
4048_586_118 <sup>2</sup>	F: CTGGGATTTGCTCGGAGAA R: GGTTTTCTGGCTTAGCTCATG	GP
25749_180_43	F: TCCACAGTGTAGCCACCAAC R: CGTCAGTGGCAAAGAGGAGT	IDS

<sup>1</sup>SNP en secuencia mitocondrial; <sup>2</sup>No amplificó; <sup>3</sup>locus con tres alelos

## Anexo F. Secuencia de oligos para preparación de librerías de GT-seq

Secuencia de oligos i5 e i7 para indexado de productos de PCR con doble código de barras para identificación de cada individuo analizado. En rojo, secuencia de código de barras.

ID oligo	Secuencia
i5Tag_AAACGG	AATGATACGGCGACCACCGAGATCTACACA <b>AAACGG</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_ACTCTT	AATGATACGGCGACCACCGAGATCTACACA <b>ACTCTT</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_ATTCCG	AATGATACGGCGACCACCGAGATCTACACA <b>ATTCCG</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_CGAGGC	AATGATACGGCGACCACCGAGATCTACACA <b>CGAGGC</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_GAAATG	AATGATACGGCGACCACCGAGATCTACACA <b>GAAATG</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_GGACTT	AATGATACGGCGACCACCGAGATCTACACA <b>GGACTT</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_TAAGCT	AATGATACGGCGACCACCGAGATCTACACA <b>TAAGCT</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_TCTTCT	AATGATACGGCGACCACCGAGATCTACACA <b>TCTTCT</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_AACGTT	AATGATACGGCGACCACCGAGATCTACACA <b>AACGTT</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_ACTGGC	AATGATACGGCGACCACCGAGATCTACACA <b>ACTGGC</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_CAAAAA	AATGATACGGCGACCACCGAGATCTACACA <b>CAAAAA</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_CGCAGA	AATGATACGGCGACCACCGAGATCTACACA <b>CGCAGA</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_GAACCA	AATGATACGGCGACCACCGAGATCTACACA <b>GAACCA</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_GGCAAG	AATGATACGGCGACCACCGAGATCTACACA <b>GGCAAG</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_TAATTC	AATGATACGGCGACCACCGAGATCTACACA <b>TAATTC</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_TGAACC	AATGATACGGCGACCACCGAGATCTACACA <b>TGAACC</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_AACTGA	AATGATACGGCGACCACCGAGATCTACACA <b>AACTGA</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC
i5Tag_AGCCAT	AATGATACGGCGACCACCGAGATCTACACA <b>AGCCAT</b> CGACAGGTTCAGAGTTCTACAGTCCGACGATC

i5Tag\_CAATCG AATGATACGGCGACCACCGAGATCTACACCAATCGCGACAGGTTTCAGAGTTCTACAGTCCGACGATC  
i5Tag\_CGCGTG AATGATACGGCGACCACCGAGATCTACACCGCGTTCGACAGGTTTCAGAGTTCTACAGTCCGACGATC  
i5Tag\_GACGAC AATGATACGGCGACCACCGAGATCTACACGACGACCGACAGGTTTCAGAGTTCTACAGTCCGACGATC  
i5Tag\_GGGCGC AATGATACGGCGACCACCGAGATCTACACGGGGCCGACAGGTTTCAGAGTTCTACAGTCCGACGATC  
i5Tag\_TACACA AATGATACGGCGACCACCGAGATCTACACTACACACGACAGGTTTCAGAGTTCTACAGTCCGACGATC  
i5Tag\_TGACAA AATGATACGGCGACCACCGAGATCTACACTGACAACGACAGGTTTCAGAGTTCTACAGTCCGACGATC  
GTseq\_i7\_i01 CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i02 CAAGCAGAAGACGGCATAACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i03 CAAGCAGAAGACGGCATAACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i04 CAAGCAGAAGACGGCATAACGAGATTGGTCAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i05 CAAGCAGAAGACGGCATAACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i06 CAAGCAGAAGACGGCATAACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i07 CAAGCAGAAGACGGCATAACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT  
GTseq\_i7\_i08 CAAGCAGAAGACGGCATAACGAGATCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

---

## Anexo G. Comandos de Perl para obtención de genotipos de librerías GT-seq

Los comandos siguientes se encuentran en el archivo *GTseq\_Genotyper\_v2.pl* en un *Github* que puede ser consultado en <https://github.com/GTseq/GTseq-Pipeline> reportado en Campell *et al.* (2015). Este archivo fue modificado para que realice el conteo de alelos con un mínimo de 4× de profundidad de lectura.

```
#!/usr/bin/perl

#GTseq_Genotyper_v2.pl by Nate Campbell
#Genotype GTseq libraries from individual .fastq files
#100x faster and more accurate than the original version of the GTseq_Genotyper.pl script. This
#version makes use of the forward primer sequences for each locus.
#Requires          .csv          file          of          [Locus
Name,Allele1,Allele2,ProbeSeq1,ProbeSeq2,FWD_Primer,A1_correction,A2_correction] for all
loci as $ARGV[0] and an individual fastq file.
#Correction values for input file are optional. If correction values are not used, then allele
correction values will be zero at all loci.
#Output files contain a header line with summary information followed by locus specific data.
#Locus specific output fields are LocusName, Allele1_counts, Allele2_counts, A1/A2-ratio,
Genotype, Genotype_Class, A1_correction value, A2_correction value, On_target reads, Locus
OT_percentage, Locus On-target reads as percentage of total on-target reads
# The penultimate field is defined as: #reads beginning with forward primer sequence and
containing a probe sequence / #All reads beginning with forward primer sequence *100
# The final field is defined as: total number of on-target reads for locus / Overall on-target reads
in panel *100

use strict; use warnings;

die "usage: provide csv file of locus, alleles, and probe seqs plus individual fastq\n" unless @ARGV
== 2;

my %On_Target = ();
my %Off_Target = ();
my %F_Primer = ();
my %F_PrimerKey = ();
```

```

my %allele1name = ();
my %allele2name = ();
my %probeA1 = ();
my %probeA2 = ();
my %probeA1_RC = ();
my %probeA2_RC = ();
my %Allele1_Count = ();
my %Allele2_Count = ();
my %A1_corr = ();
my %A2_corr = ();
my $ArraySize = 0;
my $OT_Reads = 0;
my $Raw_Reads = 0;
#read in assay and allele information and push to hashes...
open(PROBES, "<$ARGV[0]>") or die "error reading $ARGV[0]\n";
while (<PROBES>) {
    chomp;
    my @info = split(/,/ , $_);
    $F_Primer{$info[0]} = substr $info[5], 0, 14;
    $F_PrimerKey{$F_Primer{$info[0]}} = $info[0];
    $allele1name{$info[0]} = $info[1];
    $allele2name{$info[0]} = $info[2];
    $probeA1{$info[0]} = $info[3];
    $probeA2{$info[0]} = $info[4];
    if (exists $info[6]) {$A1_corr{$info[0]} = $info[6]}
    else {$A1_corr{$info[0]} = 0}
    if (exists $info[7]) {$A2_corr{$info[0]} = $info[7]}
    else {$A2_corr{$info[0]} = 0}
    my $revcomp1 = reverse $info[3];
    $revcomp1 =~ tr/[ACGT]/[TGCA]/;
    $revcomp1 =~ tr/\[\]/[\]/;

```



```

$probeA1_RC{$info[0]} = $revcomp1;
my $revcomp2 = reverse $info[4];
$revcomp2 =~ tr/[ACGT]/[TGCA]/;
$revcomp2 =~ tr/\[\^\[\]/;
$probeA2_RC{$info[0]} = $revcomp2;
#Initialize Allele counts at zero...
$Allele1_Count{$info[0]} = 0;
$Allele2_Count{$info[0]} = 0;
$On_Target{$info[0]} = 0;
$Off_Target{$info[0]} = 0;
$ArraySize++;
    }
close PROBES;
#Count alleles...
open(FASTQ, "<$ARGV[1]>" or die "error reading $ARGV[1]\n";
while (<FASTQ>) {
    my $R1_ID1 = $_;
    my $R1_seq = <FASTQ>;
    my $R1_ID2 = <FASTQ>;
    my $R1_qual = <FASTQ>;
    chomp ($R1_seq);
    my $FP_seq = substr $R1_seq, 0, 14;
    $Raw_Reads++;
    if(exists $F_PrimerKey{$FP_seq}) {
        my $target = $F_PrimerKey{$FP_seq};
        if ($R1_seq =~ m/$probeA1{$target}|$probeA1_RC{$target}/){
            $Allele1_Count{$target}++; $On_Target{$target}++; $OT_Reads++;
        }
        elsif ($R1_seq =~ m/$probeA2{$target}|$probeA2_RC{$target}/){
            $Allele2_Count{$target}++; $On_Target{$target}++; $OT_Reads++;
        }
    }
}

```

```

        else {$Off_Target{$target}++}
    }
}
close FASTQ;
my $OT_Percentage = $OT_Reads/$Raw_Reads * 100;
$OT_Percentage = sprintf("%.1f", $OT_Percentage);
# print header line...
print          "$ARGV[1],Raw-Reads:$Raw_Reads,On-Target          reads:$OT_Reads,%On-
Target:$OT_Percentage\n";
foreach my $loci (sort keys %F_Primer){
    my $A1fix = 0;
    my $A2fix = 0;
    my $sum_xy = $Allele1_Count{$loci} + $Allele2_Count{$loci};
    $Allele1_Count{$loci} = $Allele1_Count{$loci} - ($sum_xy / 4 * $A1_corr{$loci});
    if ($Allele1_Count{$loci} < 0) {$Allele1_Count{$loci} = 0}
    $Allele2_Count{$loci} = $Allele2_Count{$loci} - ($sum_xy / 4 * $A2_corr{$loci});
    if ($Allele2_Count{$loci} < 0) {$Allele2_Count{$loci} = 0}
    $Allele1_Count{$loci} = int ( $Allele1_Count{$loci} );
    $Allele2_Count{$loci} = int ( $Allele2_Count{$loci} );
    my $geno = "00"; #Initialize genotype variable at "00"...
    my $genoclass = "NA"; #Initialize genotype classification at "NA"...
#Fix allele counts to non-zero number for division ratio calculation...
    if ($Allele1_Count{$loci} == 0) {$A1fix = 0.1}
    else {$A1fix = $Allele1_Count{$loci}}
    if ($Allele2_Count{$loci} == 0) {$A2fix = 0.1}
    else {$A2fix = $Allele2_Count{$loci}}
    my $ratio = $A1fix/$A2fix;
    $ratio = sprintf("%.3f", $ratio);

    if ($Allele1_Count{$loci} + $Allele2_Count{$loci} < 4) {$geno = "00"; $genoclass = "NA";}
#Set genotypes of low allele count loci to "00"

```

```

    elif ($ratio >= 10) {$geno = "$allele1name{$loci}$allele1name{$loci}"; $genoclass =
"A1HOM";} #Allele1 Homozygotes
    elif ($ratio <= 0.1) {$geno = "$allele2name{$loci}$allele2name{$loci}"; $genoclass =
"A2HOM";} #Allele2 Homozygotes
    elif ($ratio <= 0.2) {$geno = "00"; $genoclass = "NA";} #In-betweeners
    elif ($ratio <= 5) {$geno = "$allele1name{$loci}$allele2name{$loci}"; $genoclass = "HET";}
#Heterozygotes
    if($Off_Target{$loci} == 0) {$Off_Target{$loci} = 0.01}
    my $On_Target_Per = ($On_Target{$loci}/($Off_Target{$loci} + $On_Target{$loci})) * 100;
    my $Per_of_AllOTreads = $On_Target{$loci}/$OT_Reads * 100;
    $On_Target_Per = sprintf("%.1f", $On_Target_Per);
    $Per_of_AllOTreads = sprintf("%.3f", $Per_of_AllOTreads);
    print
"$loci,$allele1name{$loci}=$Allele1_Count{$loci},$allele2name{$loci}=$Allele2_Count{$loci},$ra
tio,$geno,$genoclass,$A1_corr{$loci},$A2_corr{$loci},$On_Target{$loci},$On_Target_Per,$Per_
of_AllOTreads\n";
    }

```

## Anexo H. Archivo de entrada para obtención de genotipos en Perl de librerías GT-seq

Los comandos en Perl para la obtención de genotipos (Anexo G) requieren de un archivo csv que contiene en la primera columna el nombre del SNP, la segunda y tercer columna corresponden a los alelos esperados, la cuarta y quinta columna es la secuencia del amplicon esperado para ambos alelos (los dos nucleótidos en corchetes indican que cualquiera de las dos bases esta presentes en esa posición de la secuencia), la sexta columna es la secuencia del oligo en dirección sentido, la séptima y octava columna corresponden a valores de corrección para ambos alelos, estas dos columnas son opcionales por lo que pueden ir en 0. A continuación la información contenida en el archivo para el panel de 118 SNPs para GT-seq. En el archivo no deben llevar título las columnas. Se excluyó el SNP 4048\_586\_118 (falla en la amplificación) y 10471\_66\_124 (presencia de tres alelos).

Datos de panel de SNPs para su obtención de genotipos en Perl. (A1) alelo 1; (A2) alelo 2; (Seq1) secuencia esperada para alelo 1; (Seq2) secuencia esperada para alelo 2; (PrimerF) secuencia sentido de oligo.

SNP ID	A1	A2	Seq1	Seq2	PrimerF
>10168_152_75	T	A	TGAC[AG]GTCT[CT]GTACTT	TGAC[AG]GTCA[CT]GTACTT	ACCCAAGCTGCCTCAAAAGT
>10805_123_53	A	G	CAA[CT]AGCT[AG]TA[AG]CATCGA	CAG[CT]AGCT[AG]TA[AG]CATCGA	ATGCCCGAAGTTTGGAGGAA
>10925_305_127	T	C	GATAAGACTGTGTTCC	GATAAGACTGTGCTCC	CACCGGTAAGAGCATGTCCA
>11208_39_117	A	G	AATCAAGTAAGT	AATCAGGTAAGT	CAGATGCCATTCCAGGACGA
>11321_70_27	G	A	TAGGTGACCCGT	TAGGTAACCCGT	CTCCCAGATTTGAGGAGCCG
>11392_361_39	C	T	GTCCAGTCACTG	GTCCAGTACTG	ACCAGGAATCACTTTGCTTGC
>11550_33_63	G	A	AGGGTTAAGGG[GT]TCTCC	AGGGTTAAGAG[GT]TCTCC	CTTTCGCTGGGAGGACCTTT
>12227_85_38	G	A	CC[TC]CTGTGAAGC	CC[TC]CTATGAAGC	CACGCACATGCACATCCAC
>12314_124_128	C	T	AGCGGCGCTCCT	AGCGGTGCTCCT	TTTAGTGTGTGACGCCCTCG
>12458_182_40	G	A	TCA[GA]CGCAAGCG	TCA[GA]CACAAGCG	TACCACCTCGAGCCTCTGG

>12523_426_47	C	T	GC[GA]AGCACGGAGAAGG	GT[GA]AGCACGGAGAAGG	CCCCAATCCCCTTCTCTG
>13007_59_30	G	C	TCCTTGTGA[TC]ATAAT	TCCTTCTGA[TC]ATAAT	TCGTCTTCTGTTGTACATCGTTT
>13056_69_23	C	T	GCAGCC[GA]GGCAAT	GCAGCT[GA]GGCAAT	ACAGGCGGTAAGTATGCAGC
>14032_58_121	C	T	TTGGGCGTGTGA	TTGGGTGTGTGA	TGAGCCAATGAGTCAGCGAC
>14061_31_76	T	G	CCCCTGCGCGCGACGG	CCCCGGCGCGCGACGG	TGAGGATGAGCACCTTGAGC
>14108_1170_97	G	C	GCCTT[GA]CTTCACGCCACT	CCCTT[GA]CTTCACGCCACT	CAAAGAAGGTTGGGGCAAGC
>14538_61_117	G	A	TCGCGGGA[TC]CGAGC	TCGCGAGA[TC]CGAGC	CAGGTCATAGCCCTTCCACG
>14623_62_83	A	T	CCTCGTGATCGGTGGC	CCTCGTGTTCGGTGGC	AACGACAAGACGCGGTTTTG
>14681_61_74	T	C	TGAGTTAAATTG	TGAGTCAAATTG	ACATTTCACTTGTGTACCAGGGA
>14885_52_87	C	T	GCCTCTACCACCAGCCTT	GCCTCTACTACCAGCCTT	CCCTGTGTCTGGAGGTTTGT
>15073_47_111	C	A	AGGGTCGCCCTC	AGGGTAGCCCTC	CAGGTTTTGCCTCCTCCTCC
>15250_39_38	T	C	CCTCGTTTACCCTAAT	CCTCGTTTACCCTAAT	ATCCGGAGCAGAGTTTACCC
>16243_36_86	C	T	CCCAG[TA]G[TA]TGC[TC][CT]CACA	CCCAG[TA]G[TA]TGT[TC][CT]CACA	TCAGCAAAGTCAAGGTCGGT
>16534_42_89	T	A	T[AG]CTTTACAGTGGCA	T[AG]CTATACAGTGGCA	CTCCGGCAACCATCCAATCT
>16648_417_58	G	A	GAGG[TC]CTTTTAGAG	GAAG[TC]CTTTTAGAG	TCACGGCATCTGTGAAGTCA
>16778_168_114	A	G	AAATATCCCGC	AAATGTCCCGC	TGTCTCCCGCGAGTCTCATA
>16885_30_29	A	T	CTCTAAGTTGCA	CTCTATGTTGCA	GTACCCAGGTCGAGCCGC
>17577_54_100	G	A	GAGAGGGAAGGC	GAGAGAGAAGGC	AGAGAGAGGGGAGTGCAGGAG
>17599_44_83	C	T	TG[CT]GCCCCGCTGCTGA	TG[CT]GCTCGCCTGCTGA	TGTCGGAGGGGTTATGAGGT
>17752_60_83	T	C	TCAGGTTCACTG	TCAGGCTCACTG	CCATACAGCACGGGGTTGAT
>17883_117_37	T	A	TTGCATGGGACA	TTGCAAGGGACA	CGTGAGGGTTGGGTTCTCTC
>17910_36_120	T	A	CGCGATGTCCCC	CGCGAAGTCCCC	CGCTATTGTTGGGGGTCACT
>18219_42_111	G	A	CGCGTGCGTTCT	CGCGTACGTTCT	CGACGTCGTTTGTGAGGGAT

>18433_37_100	A	G	GTTGCTA[AG]TCTAGGCG	GTTGCTG[AG]TCTAGGCG	TCAGAGTCCCATCCCCTT
>19183_297_42	T	C	CAAAGTTGTATG	CAAAGTTGTACG	GGACAGCTCCCTGGTCAAAG
>19292_52_31	C	A	ATGTGCTGCTAG	ATGTGATGCTAG	TCTGTGCACTGCCCTTCG
>19379_31_100	T	G	TAATGA[AG]GATTTGAG	TAAGGA[AG]GATTTGAG	TCTTTGCCTCTCTGAGCAGC
>1943_140_84	T	C	TCGTGTACAGCG	TCGTGCACAGCG	CCGCGATTTATTTGGCGTGA
>19529_54_78	G	T	CCGGGCCACACACTGTGCC	CCTGGCCACACACTGTGCC	TAGGAGATGCTAGGGCCGAG
>19680_80_76	C	T	GTCCTCTTCAAGATTA	GTCCTTTTCAAGATTA	ACCTTCTCTCGGCACCATTG
>19765_60_72	G	T	ACTTAGAGATCC	ACTTATAGATCC	GGAATGCTGTACAGTGAAAGG
>20437_30_59	A	G	G[CT]CATAGGCATGG	G[CT]CATGGGCATGG	GAACAGCTTCTGGGAGCCAT
>20522_88_45	G	T	CGTTTGAAGGCT	CGTTTTAAGGCT	AGCCTCGTTAGCAGACGTTT
>20566_45_29	C	T	CTCGCCTCTCGG	CTCGCTTCTCGG	AGAGTATCCTGCGACTCCCC
>20581_53_46	G	A	CGCGAGGGCGTC	CGCGAGGACGTC	CACGGGTCGGACTCGAAC
>207_101_94	C	T	CAACCGAC[AG]TGC	CAACCGAT[AG]TGC	TGGGCATAAAGGTTCTTATCAACC
>21052_66_37	T	C	TCGTTTGATCCA	TCGTTGATCCA	TGCCCCAGTGTTACTCCCTA
>2145_32_41	A	G	ACGGCG[AG][CG]AATGGTCTGAG	ACGGCG[AG][CG]GATGGTCTGAG	AATCAGCAGGGCGCCGAG
>21571_174_67	A	G	CTGC[CT]ATAGCCATAT	CTGC[CT]GTAGCCATAT	TGTTGACGGTCCTTCCTGTG
>21988_37_80	C	T	CCCTTCCCTTCT	CCCTTCCCTTCT	TGAACTCACCAATACCCCC
>22254_93_66	C	T	GTAACCTTGATATT	GTAACCTTGATATT	CATCAGCAAAGTCAAGGTCAGT
>22476_60_84	C	T	CGAGACACGAAG	CGAGATACGAAG	GCCAAGAAGGTGTTGGGAGA
>22879_167_52	C	T	CAGCTCCACCTC	CAGCTTACCTC	GAGAACACCTAACCGCAGCT
>23048_265_32	C	T	TCACACTCTCACTGTC	TCACACTTCTCACTGTC	TGGTAGGATCTAAAAGTGTCACAC
>23179_194_57	A	T	TTGGAATAAAGAG[CT]	TTGGAATAAAGTG[CT]	GCCTCTCAAGCTGACAGTACA
>23241_32_89	A	G	ATCATATTGCAA	ATCATGTTGCAA	TCCATCCTCTCGTTCGCTTG

>2346_202_39	T	G	CTCGAGGACTTG	CTCGAGGACGTG	AGACG TTCAGATGGCTCGAG
>2358_214_66	C	G	CGGCTTTCCTGGGCC	CGGCTTTCGTGGGCC	GAGCGCATTAACTCGTCAGC
>23612_43_30	A	T	TGCGAATCGGTT	TGCGATTCTGGTT	CGTGATAAAAGAACTTTCTATTTTGCG
>23746_40_91	G	T	TGTTTCGCAACGG	TGTTCTCAACGG	ATTTCCATGACTCACGCCGT
>23987_42_95	G	A	TTGTTGTGATCA	TTGTTATGATCA	GCTTCTCCTTGTCTGTGCT
>2449_63_109	G	A	GCACATAGGCATATCTAC	GCACATAGACATATCTAC	TGTCTGGTTTTGCCACACT
>25084_128_106	C	T	CTTTTCAAGTGCCTTT	CTTTTCAAGTGTCTTT	ATGCCGGGCCAGAATAATGT
>2539_112_25	G	A	GAACCCACCCG[TC][CT]A	GAACCCACCCA[TC][CT]A	TGATT CAGTTGCCGAACCCA
>25396_159_54	T	C	ATCTTTTCAAGG	ATCTTCTCAAGG	TCTGTGTGCAAAGACATTTAACAAA
>25452_39_86	A	G	TATGAAAGTATT	TATGAGAGTATT	CTTGAGACGCTCTCGCTAGG
>2553_114_43	T	C	GAGCACCTTCCAAG	GAGCACCTCCAAG	CCTCCATCACCTCATCAGC
>26271_31_118	G	A	GAAGGAATCGTCGGGACT	GAAGGAATCATCGGGACT	GCTTGAGGAACTGACGGGAA
>26358_32_62	T	C	GAGCCGACCGTGAT	GAGCCGACCGCGAT	CAACAGAGGGCCCGAATCG
>26562_49_27	C	T	GACGTGGGCGAAGT	GATGTGGGCGAAGT	TCAGAGGAACCGGGTAGTGA
>26659_141_123	T	C	CAAAGATTTTCTTGATGT	CAAAGATTTTCTTGATGT	TGAGGGGTTGCGCAATACTT
>2849_118_101	C	A	TGCTAAATCTCAACTGG	TGCTAAATATCAACTGG	TCAAGGATCCCATTACCATGACT
>287_30_81	A	G	TTTGAACGGTGA	TTTGAGCGGTGA	CGACTCCCTCCCTCTTTGA
>2886_253_38	T	G	CCGGATGTGTGA	CCGGAGGTGTGA	CAGGCTGTATGGCAGGGTTCGC
>2933_38_119	C	T	AGGAGCTCAGGG	AGGAGTTCAGGG	TCTGAACACCAACTTGGGCA
>32_179_24	G	A	AAGCAGC[GT]AGG	AAGCAAC[GT]AGG	CGCCGCACGTAATAAAGCA
>3213_36_42	T	G	TTAACTTAGCCT	TTAACGTAGCCT	GGAGTCAGTGTGTGGGGAAT
>3467_56_72	A	G	CAGGGATC[CT]CGCGTTC	CAGGGGTC[CT]CGCGTTC	TGGTCTTCTGGGTGAGTCA
>3494_140_31	C	T	CATCACCGCCCGAACA	CATCATCGCCCGAACA	GGGACACCTGAGCCACATC

>3960_43_46	T	C	GTCGCTGGAAGT	GTCGCCGGAAGT	TTCTCTCCACATCCTGTGCGC
>4084_41_118	C	T	TCGAGTTCG[GA]CCTCGAC	TCGAGTTTG[GA]CCTCGAC	CAACCCGTTTGCCGCCAACG
>4335_52_100	G	A	TTCTGTGGC[TC]ACCAG	TCTGTAGC[TC]ACCAG	TCTGGCACTCTGCGGGAGGC
>4396_49_88	A	T	GCACTATCCTCT	GCACTTTCCTCT	GTGCCCTGTTTCTCTCCCTC
>4424_107_115	A	G	CC[GA]TGAG[GA]GGGAATCCG	CC[GA]TGGG[GA]GGGAATCCG	GGTCGCTTCAAGTGCCATT
>4532_52_56	A	C	CCCAA[TC]ATACAGA	CCCAA[TC]CTACAGA	CAGCCTCGCCTCTCACATAA
>5123_49_69	A	C	CCGTACAGAAC	CCGTCCCAGAAC	GACCTCCAGCAGCACAACAA
>513_155_122	G	C	GATAAGGTGTGTGGGAAA	GATAAGGTCTGTGGGAAA	ATAGGCCCTGTCTTGCGTTT
>5197_37_101	C	A	GGCGACGCGATA	GGCGAAGCGATA	CGTCCTCCTTGTGGGTGAAA
>5199_109_121	G	T	GCGGAGAAATGA	GCGGATAAATGA	CAGGTCACCCTCCTTCACAC
>5234_134_32	A	C	ATTAAAG[AG]A	ATTACAG[AG]A	GGGGTCGGCAATGTGTACAA
>5251_41_29	C	T	CACGTCG[AG]CGACATGTAT	CACGTCG[AG]TGACATGTAT	TGAAAGCTTCTTCGGAAACACG
>5309_122_48	C	A	CGAAGGACGTGAGGCA	CGAAGGAAGTGAGGCA	GCTGATCGGAAGCTTCTCGA
>5477_81_39	A	G	CGTA[AG]A[TC]TCACAACCTT	CGTG[AG]A[TC]TCACAACCTT	GTGATCACACTGGTACATGATATTTGT
>5500_31_87	A	G	GTGCGAGATCAC	GTGCGGGATCAC	CGAAGGTGAGAGTGAGAGCG
>5574_49_23	C	T	TCCGCCGAGGTC	TCCGCTGAGGTC	GCGAGATGGACCGAGGTTC
>5777_55_112	A	G	ATTGTACCAATA	ATTGTGCCAATA	TGCCCTTGTTGGAGAGCATT
>5830_30_82	T	C	AGACGTGGATAC	AGACGCGGATAC	TCTTTGGGAAAGGAGTCGCC
>5861_232_119	G	A	TACGTGTTGGGC	TACGTATTGGGC	AAGTTCTCGAAAGAAGGCGG
>6175_66_115	C	A	CGAGGCGAA[GA]AA	CGAGGAGAA[GA]AA	TTCGCAGTTTAAAATATTGGAAGCA
>6189_63_60	G	C	TCGGGGCGTCGG	TCGGGCCGTCGG	CAAGATCGCCGAGTTCTCGA
>6483_44_60	G	A	ATTGAGAGTCGCCTGG	ATTGAAAGTCGCCTGG	CAGTCTCAGAGTCGCCAAGG
>6602_50_98	A	G	GTAGAAGACTGT	GTAGAGGACTGT	GAGTCTTATCCGGGTGCGAG



>6795_169_62	G	C	GAACGGGAAAGC	GAACGCGAAAGC	GCGTGGAAATGGAGGGGAAA
>7042_73_65	C	G	GCTGCCTCGGCG	GCTGCGTCGGCG	ATATGTCCGCGGGCTTTGAT
>7087_47_70	G	A	TCCGAGGGCAAC	TCCGAAGGCAAC	GACCTCCACCCCTACACCC
>7257_133_89	G	C	AATTTGAACTAT	AATTTCAACTAT	GGATCCTTTTTAAAACCCAGTGCT
>7515_119_39	C	T	TCACGCCAGACC	TCACGTCAGACC	TGTAAGTGTGCAGTGGACGC
>759_312_40	T	C	TGTTCTCGTAAT	TGTTCCCCTAAT	AACCCACGTCACCATGAACA
>7606_31_72	G	T	T[AG]CCAGGCTTG[TC]	T[AG]CCATGCTTG[TC]	GCACTTGCATGGTAAAATCATTG
>8274_40_59	T	C	GCACTTTGGCCACCCACCGCA	GCACCTTGGCCACCCACCGCA	GGCCACGATGATCATGACCA
>8953_32_69	T	C	TTCT[CT]TCTCTTTCTC	TTCC[CT]TCTCTTTCTC	TGCGTTGCATATATCATTCCCA
>9223_34_81	A	G	AAG[GA]CCAG[CT]GTCAAC	AAG[GA]CCGG[CT]GTCAAC	AGTTATCTGCGTGGGACAGC
>953_97_86	A	C	CGCTAAAGGAAG	CGCTACAGGAAG	AGGCGATGAGGAGGGAGTAG
>9717_50_71	G	A	AGTCTGCCCAAG	AGTCTACCCAAG	ACGGAGTGAAATAGTACATTTTGAATG
>9814_36_52	T	C	TGCGTTGGATAA	TGCGTCGGATAA	GGTTTGTAGAAAATAATCATCGTTGCG
>25749_180_43	C	T	TGCGGCGCCGA	TGCGGTGCCGA	TCCACAGTGTAGCCACCAAC

---

## Anexo I. Costo de genotipificación de panel de 118 SNPs con GT-seq

El costo aproximado por muestra para la obtención de genotipos por GT-seq para un mínimo de 192 muestras sería de 10.13 USD, y de 13.91 USD incluyendo extracción de ADN. Sin embargo, si se analiza un mayor número de muestras, por ejemplo, 768 muestras en una celda de secuenciación MiSeq Reagent Micro Kit v2 (300 ciclos y 4 M de lecturas; 690 USD) el costo se reduciría a 8.45 USD, y a 12.23 USD incluyendo la extracción de ADN.

Costo de genotipificación por muestra y placa de panel de 118 SNPs por GT-seq, incluyendo extracción de ADN con kit Qiagen

N° Catálogo	Reactivo o material	Costo presentación (USD) <sup>1</sup>	Costo por unidad (USD) <sup>1</sup>	Unidades por placa	Costo por placa (USD) <sup>1</sup>	Costo por muestra (USD) <sup>1</sup>
<b>Extracción de ADN</b>						
69506	kit DNeasy blood and tissue de Qiagen (250 reacciones)	946	3.78	96 reacciones	363.26	3.78
<b>Preparación de librerías: PCR1</b>						
206152	QIAGEN Multiplex PCR Plus Kit (100)	340	0.14	336 µl	45.696	0.48
IDT oPools	236 primers en Pool (50 pmol por oligo; 400 µl)	1014.57	2.54	100 µl	253.64	2.64
<b>Preparación de librerías: PCR2</b>						
206152	QIAGEN Multiplex PCR Plus Kit (100)	340	0.14	480 µl	65.28	0.68
T4oligo	Primer i5 (100 nmol) con purificación adicional (HPLC) 10,000 µl	85.71	0.01	96 µl	0.82	0.01

T4oligo	Primers i7 (100 nmol) con purificación adicional (HPLC) 10,000 µl	85.71	0.01	96 µl	0.82	0.01
<b>Preparación de librerías: Normalización</b>						
A1051001	Sequal prep normalization plate (10 placas)	2175.36	217.54	1 placa	217.536	2.27
<b>Preparación de librerías: Purificación Perlas magnéticas</b>						
A63881	AMPure XP beads Beckman coulter (60 ml)	805.11	13.42	2.4 ml	32.20	0.34
<b>Preparación de librerías: Cuantificación Qubit</b>						
Q32854	Qubit dsDNA HS Assay Kit Termofisher (500 reacciones)	392.43	0.78	2 reacciones	1.57	0.02
PCR-05-C/Q32856/	Tubos 0.5 ml Axygen claros de pared delgada (1000 tubos)	116.20	0.12	2 tubos	0.23	0.001
<b>Secuenciación</b>						
MS-103-1001	MiSeq Reagent Nano v2 (300 ciclos) (1 M de lecturas)	491.87		0.5 celdas	245.94	2.56
FC-110-3001	Phix v3 control kit v3 (10 µl)	341.04	34.10	1 µl	34.10	0.36
<b>Consumibles generales</b>						
	Puntas para pipeta, combitips para pipeta de repetición, tubos, placa para PCR, alcohol etílico, guantes.				73.05	0.76

---

<sup>1</sup>Considerando tipo de cambio de \$17.50 pesos=1 USD