



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS VETERINARIAS Y PECUARIAS
ESCUELA DE CIENCIAS VETERINARIAS

Evaluación de la precisión de imputación de genotipos utilizando diferentes densidades para un panel de SNP (*single nucleotide polymorphisms*) de Cobia (*Rachycentron canadum*)

Sebastián Elías Zavala Marín

Proyecto de Memoria para optar al
Título Profesional de Médico
Veterinario
Departamento de Medicina
Preventiva Animal

PROFESOR GUÍA: JOSÉ MANUEL YÁÑEZ LÓPEZ
UNIVERSIDAD DE CHILE

SANTIAGO, CHILE
AÑO 2022



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS VETERINARIAS Y PECUARIAS
ESCUELA DE CIENCIAS VETERINARIAS

Evaluación de la precisión de imputación de genotipos utilizando diferentes densidades para un panel de SNP (*single nucleotide polymorphisms*) de Cobia (*Rachycentron canadum*)

Sebastián Elías Zavala Marín

Proyecto de Memoria para optar
al Título Profesional de Médico
Veterinario
Departamento de Medicina
Preventiva Animal

Nota Final:
Profesor Guía: José Manuel Yáñez
Profesor Corrector: Pedro Smith
Profesor Corrector: Héctor Uribe

PROFESOR GUÍA: JOSÉ MANUEL YÁÑEZ LÓPEZ
UNIVERSIDAD DE CHILE

SANTIAGO, CHILE
2022

*“A mis padres, María y Manuel, desde el fondo de mi corazón,
gracias por todo, los amo”*

AGRADECIMIENTOS

Un abrazo y gratitud eterna a mi familia, mi mamá, mi papá, a mis hermanos Mao y Fafo, a mis cuñados Pablo y Pame, gracias a todos, por apoyarme todo este tiempo, los amo y los amo por siempre, son mi familia y mi tesoro. A mis sobrinas Bea y Juli por hacerme un tío muy feliz y darme esperanza de seguir.

A Kota por ser mi amiga eterna, sin ti el mundo sería un mundo triste y oscuro, te amo mi Kotita hermosa. Juan Pablo y Omar, aunque nuestros caminos se distanciaron siempre los recuerdo muy cerca de mí.

A mi Aniwi por ser mi amiga fiel todos estos años y darle color a mi mundo que pensaba que era gris cuando contigo todo es colorido como un arcoíris después de la lluvia.

A mis colegas y amigos Romi, Maira, Ariel, Dánae, Tamara, Agustiniak, RodriGOD, Baltazar, Pablo, María, Pau y Carito, todo lo que viví con ustedes no tiene precio y tomaría muchas hojas describir todo lo que siento por cada uno de ustedes, pero puedo decirles que los quiero mucho, mis amigos, mis “nakamas”, ¡deaeaea!

A Fabito, Cami, Gigi, Conny, Javi y toda la gente que conocí en internado, fueron mi impulso a seguir durante internado, los quiero mucho.

A mis profesores y guías, los doctores: Luis Ibarra (Q.E.P.D.), Juan Lazo, Javiera Cornejo, Héctor Uribe, Pedro Smith, Galia Ramírez y Raúl Alegría por confiar en mí, apoyarme durante mi carrera y por sobre todo a José Manuel Yáñez por darme una oportunidad de desarrollarme como profesional y guiarme en mis momentos más oscuros.

A mis cientos de estudiantes de Bioestadística, mis queridos pupilos que recuerdo por darles clases y sobre todo a Jeffrey mi estudiante estrella, muchas gracias por conocer la estadística conmigo.

A Diego mi compañero fiel y mi héroe, eres mi inspiración para ser una mejor versión de mí mismo, si pudieses te daría todo el mundo, porque eres mi mundo, te amo mi hamstercito, es una promesa del corazón.

“La esperanza se extenderá, mientras no nos rindamos, la esperanza nos ayuda a seguir adelante, la esperanza nos da valor, la esperanza nos hace movernos, hace que el mundo gire. La puerta abierta, por mi mano, por nuestras manos, la puerta al futuro donde la esperanza y la desesperación existen juntos”. – Makoto Naegi

Índice

AGRADECIMIENTOS	3
RESUMEN	7
ABSTRACT	8
INTRODUCCIÓN	9
REVISIÓN BIBLIOGRÁFICA	11
Situación mundial y nacional de la acuicultura	11
Diversificación de la producción acuícola.....	11
Cobia	12
Mejoramiento genético en acuicultura.....	12
Marcadores moleculares.....	13
Selección genómica.....	14
Desequilibrio de ligamiento.....	14
Uso de SNPs de alta densidad en selección genómica	15
Estrategias para la imputación	16
OBJETIVO GENERAL	18
OBJETIVOS ESPECÍFICOS.....	18
MATERIALES Y MÉTODOS	19
Muestras de cobia	19
Datos de secuencia.....	19
Control de calidad	19
Descripción genómica y estructura poblacional usando información de SNPs	20
Imputación y precisión del genotipo	20
RESULTADOS	22
Datos de secuencia y control de calidad.....	22
Descripción genómica y estructura poblacional usando información de SNPs	22
<i>Pruning</i> de desequilibrio de ligamiento de marcadores	24
Imputación y precisión del genotipo	24
DISCUSIÓN	29
CONCLUSIÓN	31
BIBLIOGRAFÍA	32

ÍNDICE DE TABLAS

Tabla 1: Número de SNPs obtenidos a través del control de calidad para <i>call rate</i> de SNPs, frecuencia de alelo menor <i>MAF</i> y Equilibrio de Hardy Weinberg realizado en PLINK v1.09	22
Tabla 2: Promedio, desviación estándar (D.E.) y mediana de los valores de <i>MAF</i> en los marcadores SNPs posterior al control de calidad; número de marcadores con $MAF < 0,05$ y porcentaje en relación al total de marcadores posterior al control de calidad.....	23
Tabla 3: Densidades de marcadores SNPs después del <i>LD pruning</i> según R^2	24
Tabla 4: Precisión de imputación como correlación promedio entre los genotipos verdaderos e imputados a distintas densidades de marcadores (2K, 1K, 0,5K y 0,1K) y su respectiva desviación estándar (DE)	25
Tabla 5: Número de marcadores con precisión de imputación mayor al 80% según densidad (2K, 1K ,0,5K y 0,1K) y su porcentaje en relación al panel de referencia (9K)	28

ÍNDICE DE FIGURAS

Figura 1: Distribución de precios de genotipado según densidad de marcadores SNPs en USD\$.....	16
Figura 2: Distribución de frecuencias de alelo menor (<i>MAF</i>) de los marcadores posterior al control de calidad	23
Figura 3: Análisis de componentes principales de los datos genotípicos de Cobia, se observa tres subpoblaciones agrupadas.....	24
Figura 4: Precisión de imputación según cromosomas para densidades 2K(A), 1K(B), C (0,5K) y D(0,1K)	27

RESUMEN

La producción de cobia (*Rachycentron canadum*) se verá ampliamente beneficiada con la implementación de un programa de mejoramiento genético. La aplicación de paneles de SNP de alta densidad permitirá implementar metodologías de mejoramiento, tales como la selección genómica (GS) y estudios de asociación del genoma completo (GWAS), los cuales pueden acelerar el progreso genético de la especie. La aplicación de paneles de SNP de alta densidad implica costos elevados de genotipado, limitando su implementación.

La imputación de genotipos permite disminuir el costo del uso de información genómica, mediante la utilización de paneles de SNP de baja densidad imputados a mayor densidad, con una precisión determinada. El presente trabajo apuntó a evaluar la precisión de imputación desde un panel de baja densidad a un panel de alta densidad bajo diferentes densidades de paneles de SNP en una población de cobia. Un total de 192 animales fueron genotipados mediante metodología ddRAD-seq, generando 8.894(9K) posterior al control de calidad. La precisión de imputación fue evaluada mediante diferentes densidades de paneles de SNPs (2K, 1K, 0,5K y 0,1K) obtenidos mediante *pruning* del desequilibrio de ligamiento del panel de 9K usando diferentes valores de R^2 (0,29, 0,13, 0,048 y 0,003). Para la imputación un esquema quíntuple de validación cruzada fue aplicado usando como grupo de referencia un 80% de los individuos y un 20% de los individuos como grupo de validación. La precisión de imputación de los genotipos se evaluó mediante la correlación entre genotipos observados e imputados, la cual fue mayor en la densidad 2K ($0,53 \pm 0,19$), seguido por 1K ($0,36 \pm 0,21$), 0,5K ($0,20 \pm 0,21$) y 0,1K ($0,06 \pm 0,16$). Las diferencias entre las precisiones de imputación según las distintas densidades pueden ser explicadas por: densidad de marcadores en los animales como referencia y la cantidad de estos, ausencia de pedigrí disponible, parentesco de la población, variabilidad genómica de los marcadores y presencia de *gaps* al genotipar con metodología ddRAD-seq. Concluyendo que la densidad de 2K permite obtener un panel de 6K con la mayor precisión de imputación, abaratando un 27% de los costos, en una producción tipo.

Palabras claves: Cobia, imputación, precisión, SNP, ddRAD-seq

ABSTRACT

Cobia (*Rachycentron canadum*) production will greatly benefit from the implementation of a genetic improvement program. The application of high-density SNP panels will allow the implementation of breeding methodologies, such as genomic selection (GS) and genome-wide association studies (GWAS), which can accelerate the genetic progress of the species. The application of high-density SNP panels implies high genotyping costs, limiting its implementation. The imputation of genotypes makes it possible to reduce the cost of using genomic information, using panels of low-density SNPs imputed at higher density, with a certain precision. The present work aimed to evaluate the imputation accuracy from a low-density panel to a high-density panel under different SNP panel densities in a cobia population. A total of 192 animals were genotyped using the ddRAD-seq methodology, generating 8,894(9K) after quality control. Imputation accuracy was assessed using different densities of SNP panels (2K, 1K, 0.5K and 0.1K) obtained by linkage disequilibrium pruning of the 9K panel using different R² values (0.29, 0.13, 0.048 and 0.003). For the imputation, a quintuple cross-validation scheme was applied using 80% of the individuals as the reference group and 20% of the individuals as the validation group. The imputation accuracy of the genotypes was evaluated by the correlation between observed and imputed genotypes, which was higher at density 2K (0.53 ± 0.19), followed by 1K (0.36 ± 0.21), 0.5K (0.20 ± 0.21) and 0.1K (0.06 ± 0.16). The differences between the imputation precisions according to the different densities can be explained by: density of markers in the reference animals and their quantity, absence of pedigree, kinship of the population, genomic variability of the markers and presence of gaps in the genotype with ddRAD-seq methodology. Concluding that the 2K density allows to obtain a 6K panel with the highest imputation precision, lowering costs by 27%, in a typical production setting.

Keywords: Cobia, imputation, precision, SNP, ddRAD-seq

INTRODUCCIÓN

La acuicultura representa una alternativa importante para la obtención de alimento con alto valor nutricional. Sin embargo, el número de especies que conforman la producción mundial, aunque es alto, concentra su mayor producción en un grupo reducido de especies (FAO, 2018). En este sentido, la diversificación en la producción acuícola constituye un importante desafío para producir alimentos bajo el contexto de una elevada presión económica, social y climática que experimenta el mundo (Harvey *et al.*, 2016).

La cobia (*Rachycentron canadum*) es una especie marina de amplia distribución geográfica, encontrándose tanto en aguas tropicales, subtropicales y templadas (Shaffer y Nakamura, 1989). Además, se ha descrito como una especie promisoría para los programas de diversificación acuícola, ya que posee alta tasa de sobrevivencia en cautiverio, rápido crecimiento, resistencia a enfermedades y una apropiada tasa de conversión alimentaria (Gutiérrez, 2017), entre otras aptitudes productivas.

Los programas de mejoramiento genético en acuicultura tienen como objetivo mejorar la eficiencia de la producción para satisfacer las demandas de los consumidores y aumentar la rentabilidad comercial. Para mejorar la eficiencia de producción de cualquier especie acuícola, es útil caracterizar la estructura genómica, las variaciones genómicas y la base genética de rasgos económicamente importantes (You *et al.*, 2020). El desarrollo de nuevas tecnologías permite obtener mayor información del genoma, y así facilitar el mejoramiento genético, a través de la identificación de una alta cantidad de marcadores moleculares asociados a genes relacionados a rasgos productivos en diversas especies (Dufflocq *et al.*, 2019). Dentro de los distintos tipos de marcadores moleculares destacan los polimorfismos de un solo nucleótido (SNP), los cuales corresponden a los marcadores de elección para asistir los programas de mejoramiento genético (Koopae y Koshkoiyeh, 2014). La aplicación de SNPs en alta densidad permite aplicar metodologías como la selección genómica (GS) y estudios de asociación del genoma completo (GWAS) (Heidaritabar *et al.*, 2015). Una desventaja de la aplicación de SNPs en alta densidad es su elevado costo de genotipado, limitando su implementación en la práctica (Heidaritabar *et al.*, 2015; Yang *et al.*, 2020). Para enfrentar esta situación, la imputación de genotipos se ha propuesto como una estrategia costo-efectiva para permitir la implementación de

información genómica en programas de mejoramiento genético a nivel comercial (Heidaritabar *et al.*, 2015; Tsai *et al.*, 2017). Esta estrategia consiste en utilizar SNPs en baja densidad, los cuales tienen un menor costo, y a partir de éstos generar genotipos en mayor densidad mediante algoritmos informáticos que permiten inferir los genotipos de los marcadores faltantes. La precisión de imputación es afectada por factores tales como el nivel de parentesco entre los animales de la población de referencia y la población de validación, genotipados con alta y baja densidad de SNPs, respectivamente, el número de animales en la población de referencia, la distribución de los SNPs en el cromosoma, la densidad de SNPs utilizada para la población de referencia y validación (Sargolzaei *et al.*, 2014; Calus *et al.*, 2014; Bolormaa *et al.*, 2015; Ventura *et al.*, 2014).

Por lo tanto, el objetivo de este estudio fue determinar la precisión de imputación de un panel de menor a un panel de mayor densidad de SNPs, mediante diferentes densidades de SNPs en una especie acuícola comercial no tradicional, como lo es la cobia.

REVISIÓN BIBLIOGRÁFICA

Situación mundial y nacional de la acuicultura

La producción total de pescado alcanzó un máximo histórico de 171 millones de toneladas en 2016, de las cuales el 88% se destinaron al consumo humano. De esta forma, el consumo per cápita alcanzó los 20,3 kg, representando un máximo histórico y poniendo de manifiesto que el sector pesquero es fundamental para alcanzar la meta de la FAO de combatir el hambre y desnutrición mundial (FAO, 2018).

En este contexto, la acuicultura aportó para el año 2016 un total de 80,0 millones de toneladas de peces comestibles, siendo el continente asiático, el responsable del 89% de la producción acuícola mundial. América representa el 4,54% de la producción total de la acuicultura, en la que Chile tiene una participación del 36,34% en relación con al resto del continente, posicionándose como el principal productor acuícola de América (FAO, 2018).

Chile cuenta con condiciones geográficas favorables para el desarrollo de la acuicultura. La actividad acuícola de mayor importancia a nivel nacional corresponde a la salmonicultura, la cual ha presentado un desarrollo sostenido durante las últimas décadas. A enero de 2020 la cosecha acumulada fue de 953.3 mil toneladas totales, donde el salmón del Atlántico (*Salmo salar*) aportó con un 73,2%, el salmón del Pacífico (*Oncorhynchus kisutch*) con un 18,5% y la trucha arcoíris (*O. mykiss*) con un 8,3% de la producción total (SUBPESCA, 2020).

Diversificación de la producción acuícola

El cultivo de peces constituye el subsector más diverso dentro de la acuicultura mundial con un total de 27 especies y grupos de especies. Sin embargo, el 50% de la producción acuícola de peces está dominado por un grupo reducido de especies: carpa herbívora (*Ctenopharyngodon idella*), carpa plateada (*Hypophthalmichthys molitrix*), carpa común (*Cyprinus carpio*), tilapia del Nilo (*Oreochromis niloticus*), carpa de cabeza grande (*Hypophthalmichthys nobilis*), *Carassius* spp. y *Catla catla* (FAO, 2018).

A partir de esta situación se plantea la necesidad de diversificar la acuicultura, la cual en general es impulsada por fuerzas del mercado, la disponibilidad reducida de una especie particular en la naturaleza, la ubicación o la disponibilidad de recursos, las

enfermedades de especies hidrobiológicas, las políticas gubernamentales, la presión social y el cambio climático (Harvey *et al.*, 2016).

Para que una especie sea una candidata interesante a ser incorporada en los programas de diversificación de la producción acuícola, se deben considerar los siguientes factores: tener un suministro confiable de semillas y supervivencia para la cosecha; ser especie eurihalina y/o euritérmica; tolerar contaminación y bajos niveles de oxígeno; provenir de niveles tróficos menores; tener una conversión de alimento rentable; tener ciclos de producción cortos; cumplir con los requisitos de bioseguridad; y ser culturalmente aceptables y reflejar las preferencias cambiantes de los consumidores (Harvey *et al.*, 2016).

Cobia

La cobia es marina y de amplia distribución mundial, encontrándose tanto en aguas tropicales, subtropicales y templadas (Shaffer y Nakamura, 1989). Esta especie es la única representante del género *Rachycentron* y de la familia *Rachycentridae* (Gutiérrez, 2017). Los ejemplares adultos pueden medir hasta 2 metros y pesar 68 kg, presentan alta tasa de conversión alimenticia, rápido crecimiento, llegando incluso a 6 kg de peso promedio durante el primer año de cultivo, es resistente a ciertas enfermedades, presenta una aceptable sobrevivencia larvaria en cautiverio, adaptabilidad a los alimentos comerciales y capacidad para desovar en condiciones naturales e inducidas (Holt *et al.*, 2007; Gutiérrez, 2017). Todas estas características biológicas convierten a la cobia en una especie atractiva para la acuicultura, incentivando su producción en varios países (Corialano y Cohelo, 2012) y excelente candidata para la diversificación en acuicultura (Mundo-Acuícola, 2015).

Para el año 2018, las estadísticas globales sobre la producción en toneladas (t) de cobia percibieron una producción total de 45.860 toneladas. China produjo el 84,7% de la producción total, seguido por Estados Unidos con un 4,7%, Panamá con un 3,8%, Viet Nam 3,3%, Taiwan 2,7%, Ecuador 0,8% y Republica Dominicana 0,1% (FAO, 2020).

Mejoramiento genético en acuicultura

La necesidad de mejorar la rentabilidad de los sistemas acuícolas puede ser apoyada mediante la selección genética de individuos con genotipos superiores para las características comerciales de interés productivo, mediante programas de mejoramiento

animal (Gjedrem, 2012). Para la selección, generalmente se consideran las características heredables y económicamente importantes por ejemplo: tasa de crecimiento, calidad de producto y la resistencia a enfermedades (Gjedrem, 2012). Sin embargo, a pesar de los beneficios que conlleva la aplicación de estos programas en la producción animal, se estima que sólo el 10% de la producción acuícola mundial está basada en poblaciones que han sido sometidas a programas de mejoramiento genético (Perez-Enriquez y Max-Aguilar, 2016)

La mejora genética está basada en la selección de peces de valor genético aditivo superior, los cuales serán reproducidos para generar la siguiente generación en la población de cultivo (Dufflocq *et al.*, 2019). Tradicionalmente el método más utilizado para la seleccionar individuos se basa en información fenotípica e información genealógica, la cual permite estimar los efectos fijos (factores y co-variables) y predecir los efectos aleatorios (valores genéticos aditivos) que afectan a las características de interés, por medio del método BLUP (*Best Linear Unbiased Prediction*) (Henderson, 1984; Dufflocq *et al.*, 2019).

Marcadores moleculares

El desarrollo de nuevas tecnologías moleculares que permiten obtener mayor cantidad y mejor calidad de información del genoma, sumado a la constante disminución en los costos de obtención de esta información, han modificado progresivamente los escenarios sobre los cuales se puede realizar el mejoramiento genético. Estos avances permiten la posibilidad de identificar una alta cantidad de marcadores moleculares asociados que pueden estar relacionados a alguna característica productiva en diversas especies (Dufflocq *et al.*, 2019). La identificación y utilización de marcadores moleculares puede proporcionar una mejora considerable a la respuesta de selección, por ejemplo, mediante selección asistida por marcadores (MAS) o selección genómica (Haldar, 2018). Dentro de la variedad de marcadores de ADN que existen, destacan los polimorfismos de un solo nucleótido (SNP), que implican la sustitución de un nucleótido por otro, es decir, el cambio de una base nitrogenada única en una secuencia de ADN (Beuzen *et al.*, 2000).

El uso de este tipo de marcador molecular es de elección gracias a su abundancia en los genomas de diversas especies y posibilidad de automatización en su análisis, el cual, además se realiza con elevada precisión (Vignal *et al.*, 2002). Debido a esto, los SNPs son

de gran utilidad para asistir las evaluaciones genéticas empleando información genómica, construcción de mapas genéticos y el mapeo de QTL (*Quantitative trait loci*) a través de GWAS (*Genome-Wide Association Studies*) (Dufflocq *et al.*, 2019), facilitando la identificación de genes responsables de la variación genética para diversos caracteres productivos en especies de interés comercial (Hayes y Goddard, 2010; Georges *et al.*, 2018). Los SNPs se han utilizado para la caracterización de poblaciones de reproductores e identificación cepas en la acuicultura (Wenne, 2018), así como también para encontrar genes candidatos de rasgos productivos y loci de rasgos cuantitativos (QTL) útiles en la acuicultura (Oyarzún *et al.*, 2013; Yáñez *et al.*, 2015), y más recientemente en la mejora genética de especies acuícolas (Georges *et al.*, 2018; Houston *et al.*, 2020).

Selección genómica

Una de las aplicaciones más importantes de los marcadores moleculares en mejoramiento genético está relacionada con su utilización en las evaluaciones genéticas, en lo que se ha denominado selección genómica (Goddard y Hayes, 2009). La selección genómica se basa en la utilización de paneles densos de SNP (es decir, miles o decenas de miles de marcadores), los cuales cubren todo el genoma, apuntando a capturar la información de todos los QTL afectando una característica particular mediante el desequilibrio de ligamiento entre éstos y al menos un marcador, permitiendo la predicción de los valores genéticos de los animales con una mayor precisión (Meuwissen *et al.*, 2001).

Entre las ventajas de la selección genómica encontramos que este método puede prescindir de los registros fenotípicos de los candidatos a selección para evaluarlos, por consiguiente, entregando una fuente de información adicional, lo cual implica mayor precisión al estimar valores genéticos, y por ende, mejor respuesta a la selección (Dufflocq *et al.*, 2019). También este método permite reducir los costos en las pruebas de progenie posibilitando la preselección de individuos de ambos sexos a edad temprana, facilitando el genotipado desde el nacimiento de los individuos y la reducción de los intervalos generacionales (Meuwissen *et al.*, 2001; Vallejo *et al.*, 2017).

Desequilibrio de ligamiento

El desequilibrio de ligamiento(LD) es una asociación no aleatoria en la aparición de alelos en dos loci(Slatkin, 2008). Para *loci* bialélicos, una de las medidas más utilizadas

para LD es R^2 (Hill y Robertson, 1968), siendo este el cuadrado del coeficiente de correlación entre dos variables indicadoras, una que representa la presencia o ausencia de un alelo particular en el primer *locus* y la otra que representa la presencia o ausencia de un determinado alelo en el segundo locus.

Uso de SNPs de alta densidad en selección genómica

El uso de SNPs en alta densidad para la selección genómica se ha vuelto habitual en los programas de mejoramiento genético animal y vegetal (Heidaritabar *et al.*, 2015). Debido a que la selección genómica explota el LD entre SNP y las variantes causales para un determinado carácter, es esperable que el incremento en la densidad de los SNPs aporte una mayor precisión de la predicción genómica (Meuwissen *et al.*, 2001; Spencer *et al.*, 2009).

Sin embargo, actualmente una mayor densidad de SNPs significa un mayor costo de genotipado, siendo una limitación clave para implementar la selección genómica en los programas de mejoramiento genético animal (Heidaritabar *et al.*, 2015; Yang *et al.*, 2020).

Actualmente los precios de genotipado para densidades 0,1K hasta 10K, van desde los 10 USD hasta los 30 USD por muestra (figura 1). Yoshida *et al.* 2019 enfatizan que el uso de estrategias de imputación puede reducir el costo de la genotipificación en al menos un 60%, dependiendo de la estrategia y número de animales analizados.

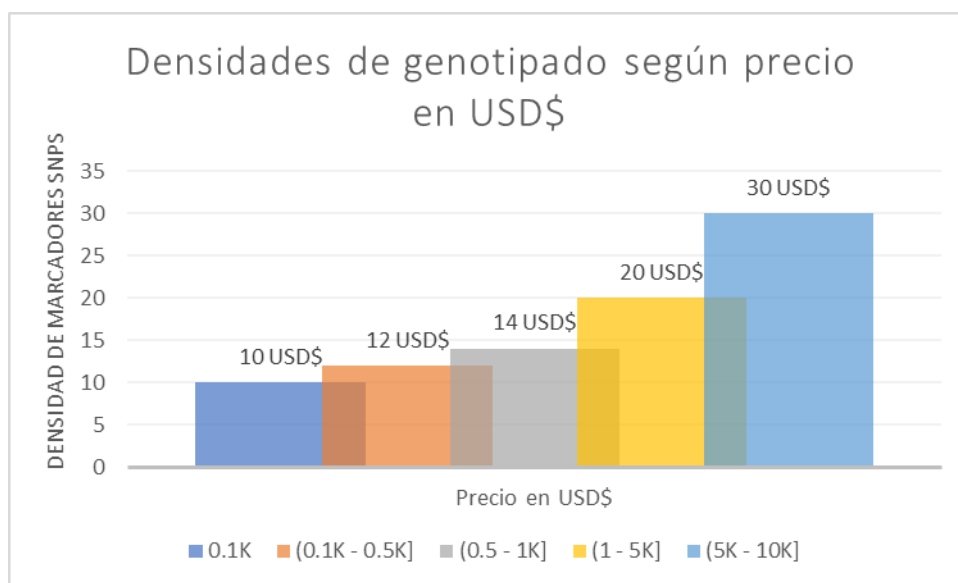


Figura 1: Distribución de precios de genotipado según densidad de marcadores SNPs en USD\$

Para superar esta restricción, los candidatos a la selección se pueden genotipar con SNPs en baja densidad, mientras que en los padres u otros animales relacionados se puede realizar la genotipificación con SNPs en mayor densidad. A través de la imputación de genotipos, se puede predecir la información de los marcadores faltantes en aquellos animales genotipados con baja densidad al de alta densidad, siendo una estrategia costo-efectiva para ser aplicada a nivel industrial (Heidaritabar *et al.*, 2015; Tsai *et al.*, 2017).

Imputación

La imputación es un método que emplea la estructura de desequilibrio de ligamiento de las regiones del genoma para inferir alelos que no han sido genotipados. Todos los métodos de imputación se basan en la Teoría de la coalescencia, que en pocas palabras, establece que se espera que un alelo de haplotipo corto (es decir, sobre una región cromosómica corta) sea compartido por muchos individuos debido a la herencia idéntica por descendencia (Wang *et al.*, 2015).

La metodología habitual de imputación se basa en tres etapas: i) inferencia de los haplotipos (*phasing*) utilizando los datos genotípicos, ii) construcción de la librería de haplotipos probables, y iii) predicción de los genotipos faltantes basándose en la información de los haplotipos de referencia (Spencer *et al.*, 2009). El método se optimiza

mediante algoritmos probabilísticos que definen qué base nitrogenada debe ser asignada a cada una de los genotipos faltantes en cada posición (Huang *et al.*, 2012).

En términos computacionales los métodos de imputación se pueden clasificar en regresión y agrupamiento (*regression and clustering*) (Wang *et al.*, 2012), modelo oculto de Márkov (HMM) y algoritmo esperanza-maximización (EM) (Browning, 2008). Actualmente existen diversos software aplicables para realizar imputación entre los cuales se encuentran: *fastPHASE*, *MaCH*, *Impute*, *Beagle*, *Mendel-Impute* y *Fimpute* (Wang *et al.*, 2015).

La precisión de imputación es afectada por varios factores, incluyendo: la estructura de la población, el número de SNPs de baja y alta densidad, el nivel de parentesco entre la población de referencia y la población de prueba, el tamaño efectivo de la población, la precisión inherente del método usado para imputar y el grado en que los marcadores están ordenados correctamente a lo largo del mapa del genoma (Hayes *et al.*, 2012; Hozé *et al.*, 2013; Uemoto *et al.*, 2015).

La cobia es una especie promisoría para incluir en los programas de diversificación acuícola por sus rasgos zootécnicos. Un programa de mejoramiento genético para esta especie requiere el desarrollo y la implementación de tecnologías genómicas, debido a la gran dificultad de realizar cruzamientos dirigidos en esta especie. Esto último permitirá mejorar la eficiencia de producción, sin embargo, el despliegue de aproximaciones genómicas debe ser equilibrado en la relación costo-beneficio, pudiendo considerar la utilización de genotipificación de marcadores de distintas densidades y la aplicación de metodología de imputación como una solución al problema del costo de genotipado

OBJETIVO GENERAL

Determinar la precisión de imputación de distintos conjuntos de SNPs de baja densidad (2K; 1K; 0,5K y 0,1K) a uno de mayor densidad (9K) en una población de cultivo de cobia.

OBJETIVOS ESPECÍFICOS

1. Realizar una caracterización genética en términos de estructura y distribución de alelo menor de un panel de mediana densidad de SNPs en cobia
2. Construcción de paneles de baja densidad de SNPs según el coeficiente de desequilibrio de ligamiento (R^2) para diferentes densidades (2K; 1K; 0,5K y 0,1K)
3. Determinar la precisión de imputación desde diferentes densidades de SNPs (2K; 1K; 0,5K y 0,1K) a una densidad intermedia (9K).

MATERIALES Y MÉTODOS

Muestras de cobia

La población de cobia que fue utilizada en el presente estudio pertenece a la empresa Open Blue, ubicada en Panamá. Se muestreo para genotipado la aleta de 192 reproductores (92♀+100♂), los cuales pertenecen a un núcleo de reproducción. La conformación del núcleo de reproducción estaba constituida por treinta estanques (3♀+2♂/estanque). Estos reproductores fueron mantenidos durante un período de 2-3 años sin ningún reemplazo hasta la nueva generación de reproductores sexualmente maduros. Open Blue facilitó la entrega de los genotipos de los 192 reproductores obtenidos mediante la técnica *ddRAD-seq* (*Double digest restriction site Associated DNA Sequencing*).

Datos de secuencia

El ADN genómico de alta calidad y peso molecular se extrajo utilizando los métodos establecidos por *Center for Aquaculture Technologies* (CAT). Las bibliotecas de secuenciación se construyeron siguiendo la digestión con la endonucleasa ApeKI y se secuenciaron utilizando una plataforma HiSeq4000. El procesamiento de los datos crudos se realizó siguiendo el *pipeline* de STACKS(Catchen *et al.*, 2011), alineando el genoma de referencia con BOWTIE2(Langmead y Salzberg, 2012) obteniendo el descubrimiento de SNPs putativos para *ddRAD-seq*.

Control de calidad

Previo al proceso de imputación, se realizó un control de calidad de los genotipos utilizando el conjunto de herramientas de análisis de asociación de genoma completo de código abierto y gratuito *PLINK* v1.09 (Purcell *et al.*, 2007) con *scripts* de lenguaje de R v4.02 (R Core Team, 2020) en *RStudio* (RStudio Team, 2020). El primer paso del control de calidad fue la remoción de la información faltante desde los genotipos empleando una tasa de llamado (*call rate*) del 90% para individuos y SNP para poder filtrar genotipos en base a la calidad del ADN presente en las muestras y la información que puedan otorgar los SNP incluidos, respectivamente. Posteriormente, se evaluó el set de datos según la frecuencia del alelo menos común (*Minor allele frequency, MAF*), entendiéndolo como la frecuencia del alelo que ocurre con menos frecuencia en una ubicación específica. Se incluyó todo marcador por sobre 0,01 de *MAF*. El siguiente filtro utilizó el equilibrio de

Hardy-Weinberg (*HWE*; $p\text{-value} < 1 \times 10^{-5}$), entendiendo *HWE* como un modelo nulo de la relación entre las frecuencias de alelos y genotipos, tanto dentro como entre generaciones, bajo supuestos de no mutación, migración, selección, apareamiento aleatorio y tamaño de población infinito (Lachance, 2016); se corrigió por Bonferroni para ajustar los valores de p de manera que se reduzca el número de falsos positivos que se informan como significativos por prueba individual (Jafari y Ansari-Pour, 2019). Se excluyeron los marcadores que se desvían del *HWE*, lo que sugiere error de genotipado (Marees *et al.*, 2018), se empleó un $p\text{-value} < 1 \times 10^{-5}$ en el filtro de *HWE*, para disminuir al mínimo la pérdida de SNPs por control de calidad.

Variabilidad genómica y estructura poblacional usando información de SNPs

Para la variabilidad genómica de la población de cobia se obtuvo la distribución de marcadores SNPs posterior al control de calidad según la frecuencia del alelo menor. Para describir la población de cobia se realizó un análisis de componentes principales (*principal component analysis*, PCA) usando los marcadores de SNP posterior al control de calidad. El análisis de componentes principales es un método estadístico que cuando es aplicado a datos genotípicos permite la detección de la estructura poblacional usando el distanciamiento genético entre individuos estimado en una matriz de parentesco genómico (Ely *et al.*, 2010). Ambos estudios se realizaron con *PLINK* v1.09 (Purcell *et al.*, 2007)

Imputación y precisión del genotipo

Se construyeron a partir del set de SNPs, paneles de diferentes densidades de SNPs (2K, 1K, 0,5K y 0,1K) obtenidos mediante *pruning* del desequilibrio de ligamiento (*LD pruning*) basándose en los valores de R^2 en *PLINK* (Purcell *et al.*, 2007). Para la imputación un esquema quíntuple de validación cruzada fue aplicado usando como grupo de referencia 80% de los individuos y un 20% de los individuos como grupo de validación. Los animales utilizados como referencia tenían la información a nivel de la máxima densidad de SNP después del control de calidad y los animales de validación tenían genotipos a densidades bajas (2K, 1K, 0,5K y 0,1K)

Para evaluar la precisión de la imputación, se empleó el cuadrado del coeficiente de correlación de Pearson entre los genotipos observados e imputados (r^2). El r^2 se estimó según el método de (Huang *et al.*, 2009):

$$r^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(g_i - \bar{g})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (g_i - \bar{g})^2}}$$

En donde para cada SNP, se enmascararon los genotipos verdaderos, se le codificó los posibles genotipos como 0, 1 o 2, que representan los recuentos posibles del alelo menor en ese SNP en la población objetivo; x_i representa el genotipo imputado para el individuo i , y sea \bar{x} el valor medio de los genotipos imputados entre los individuos; de la misma forma g_i y \bar{g} los valores análogos para los genotipos verdaderos; n representa el número de individuos en la población; posterior al cálculo, el cuadrado del coeficiente de correlación de *Pearson* fue promediado entre los SNPs para obtener una medida resumen para la población. La imputación se realizó mediante el software *FImpute* v3 (Sargolzaei *et al.*, 2014).

RESULTADOS

Datos de secuencia y control de calidad

Fueron identificadas un promedio de $5,9 \pm 2,1$ millones de lecturas por muestra. De un total de 884.876 SNPs, inicialmente, 265.252 SNPs tenían una tasa de llamado (*call rate*) del 90%. Posteriormente 9.070 marcadores tenían un MAF por sobre 0,01 y finalmente 8.894 marcadores fueron retenidos por equilibrio de Hardy-Weinberg (*HWE*; *p-value* $< 1 \times 10^{-5}$) (Tabla 1)

Filtro	Nº de SNPs
Sin control de calidad	884876
<i>Call rate</i> 90%	265252
MAF 0,01	9070
<i>HWE</i> $< 1 \times 10^{-5}$	8894

Tabla 1: Número de SNPs obtenidos a través del control de calidad para *call rate* de SNPs, frecuencia de alelo menor *MAF* y Equilibrio de Hardy Weinberg realizado en PLINK v1.09 (Purcell *et al.*, 2007)

Variabilidad genómica y estructura poblacional usando información de SNPs

Se evaluó la distribución de la frecuencia del alelo menor del conjunto de SNPs (Figura 2) después del control de calidad, el promedio de MAF de los marcadores fue de $0,11 \pm 0,1$, mediana de 0,09 y con 3.290 marcadores con un valor de MAF menor a 0,05, representando el 37% del total de marcadores posterior al control de calidad (Tabla 2).

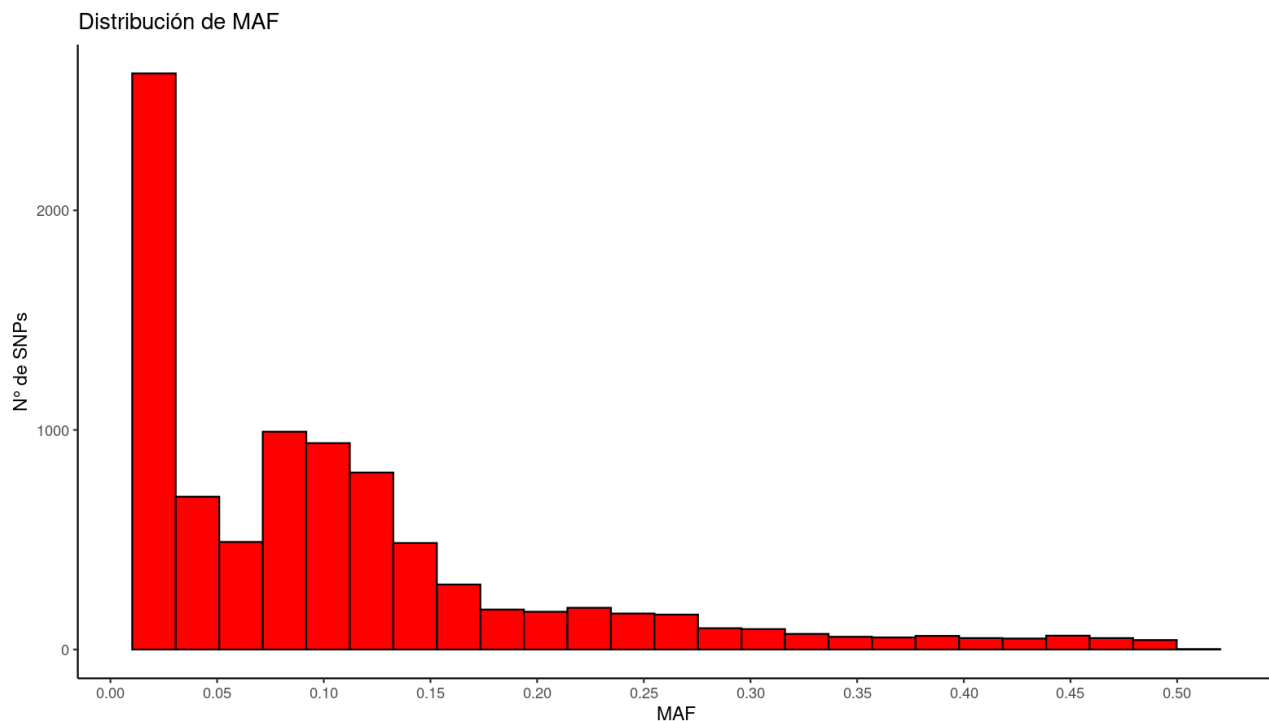


Figura 1. Distribución de frecuencias de alelo menor (MAF) de los marcadores posterior al control de calidad

Distribución de MAF			
Promedio \pm D.E.	Mediana	Nº SNPs con MAF < 0,05	%
0,11 \pm 0,1	0,09	3290	37

Tabla 2: Promedio, desviación estándar (D.E.) y mediana de los valores de MAF en los marcadores SNPs posterior al control de calidad; número de marcadores con MAF < 0,05 y porcentaje en relación al total de marcadores posterior al control de calidad.

Se observó tres poblaciones divididas en estratos en función de la variación unidimensional entre ellas después de trazar los dos primeros vectores *eigen* en el gráfico PCA (figura 3). Los dos primeros componentes principales juntos representaron el 46% de la variación genética, revelando diferentes sub-poblaciones.

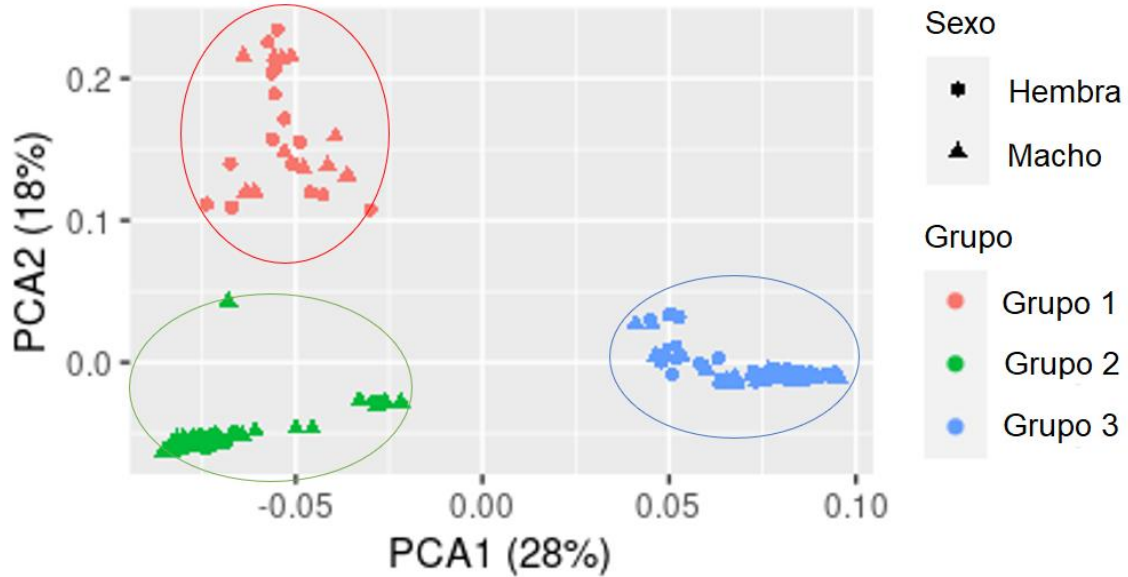


Figura 3. Análisis de componentes principales de los datos genéticos de Cobia, se observa tres subpoblaciones agrupadas

Pruning de desequilibrio de ligamiento de marcadores

Se realizó un *LD pruning* de los marcadores posterior al control de calidad, basándose en el valor de R^2 , formando cuatro grupos de densidades (2K, 1K, 0,5K y 0,1K) para cuatro valores de R^2 (0,29, 0,13, 0,048 y 0,0033) respectivamente (Tabla 3)

Densidad	Nº de SNPs	R^2
2K	2049	0,2900
1K	1004	0,1300
0,5K	505	0,0480
0,1K	100	0,0033

Tabla 3: Densidades de marcadores SNPs después del *LD pruning* según R^2

Imputación y precisión del genotipo

Se realizó la imputación mediante *Fimpute* v3 (Wang *et al.*, 2015). Se determinó la precisión del genotipo en distintas densidades, empleando como grupos de referencia 20% de los animales (N = 38) y validación 80% de los animales (N= 154). Todos los animales poseían genotipos a la mayor densidad de SNPs posterior al control de calidad (9K). Para estimar la precisión de imputación, se enmascaró diferentes densidades de SNPs en los

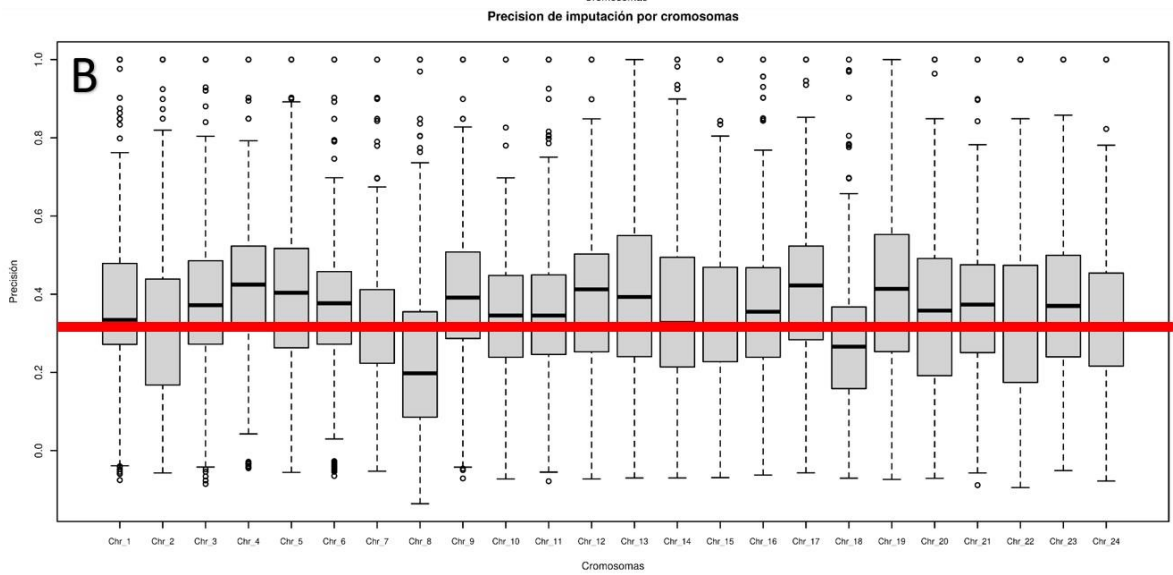
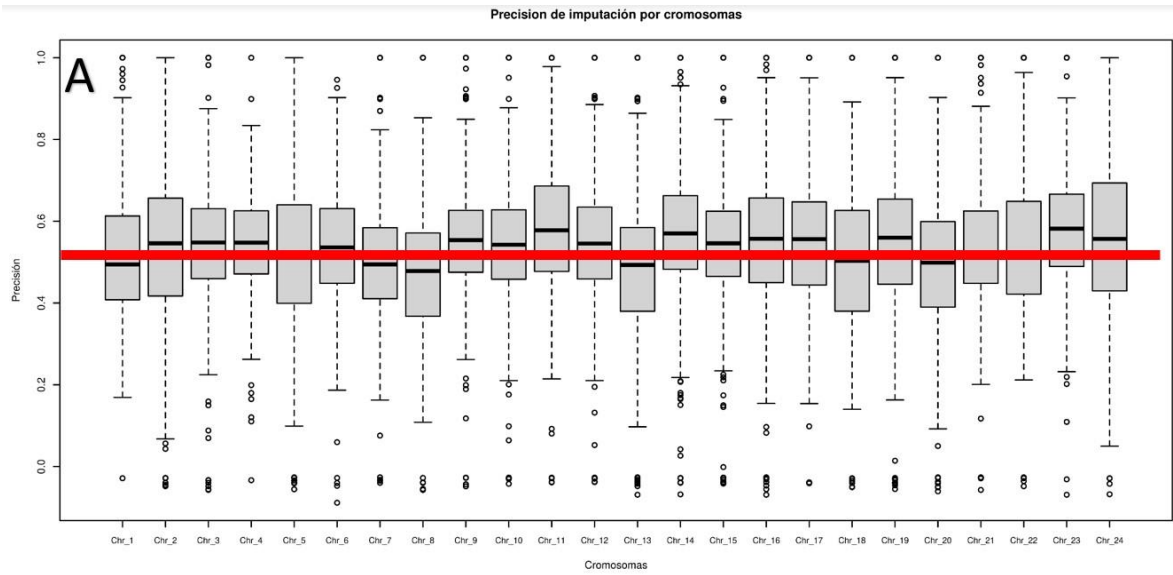
animales de validación (2K, 1K, 0,5K y 0,1K), las diferentes densidades de SNPs enmascarados fueron definidos usando un criterio de *LD pruning* (Tabla 3).

La precisión de imputación obtenida va desde $0,53 \pm 0,19$ a $0,06 \pm 0,16$ desde 2K a 0,1K, respectivamente (Tabla 4).

Densidad	Nº de SNPs	Correlación promedio \pm D.E.
2K	2049	$0,53 \pm 0,19$
1K	1004	$0,36 \pm 0,21$
0,5K	505	$0,20 \pm 0,21$
0,1K	100	$0,06 \pm 0,16$

Tabla 4: Precisión de imputación como correlación promedio entre los genotipos verdaderos e imputados a distintas densidades de marcadores (2K, 1K, 0,5K y 0,1K) y su respectiva desviación estándar (DE).

Se construyeron gráficos de caja y bigotes en (RStudio Team, 2020) mostrando la precisión de imputación según cromosomas y densidades empleadas(Figura 4)



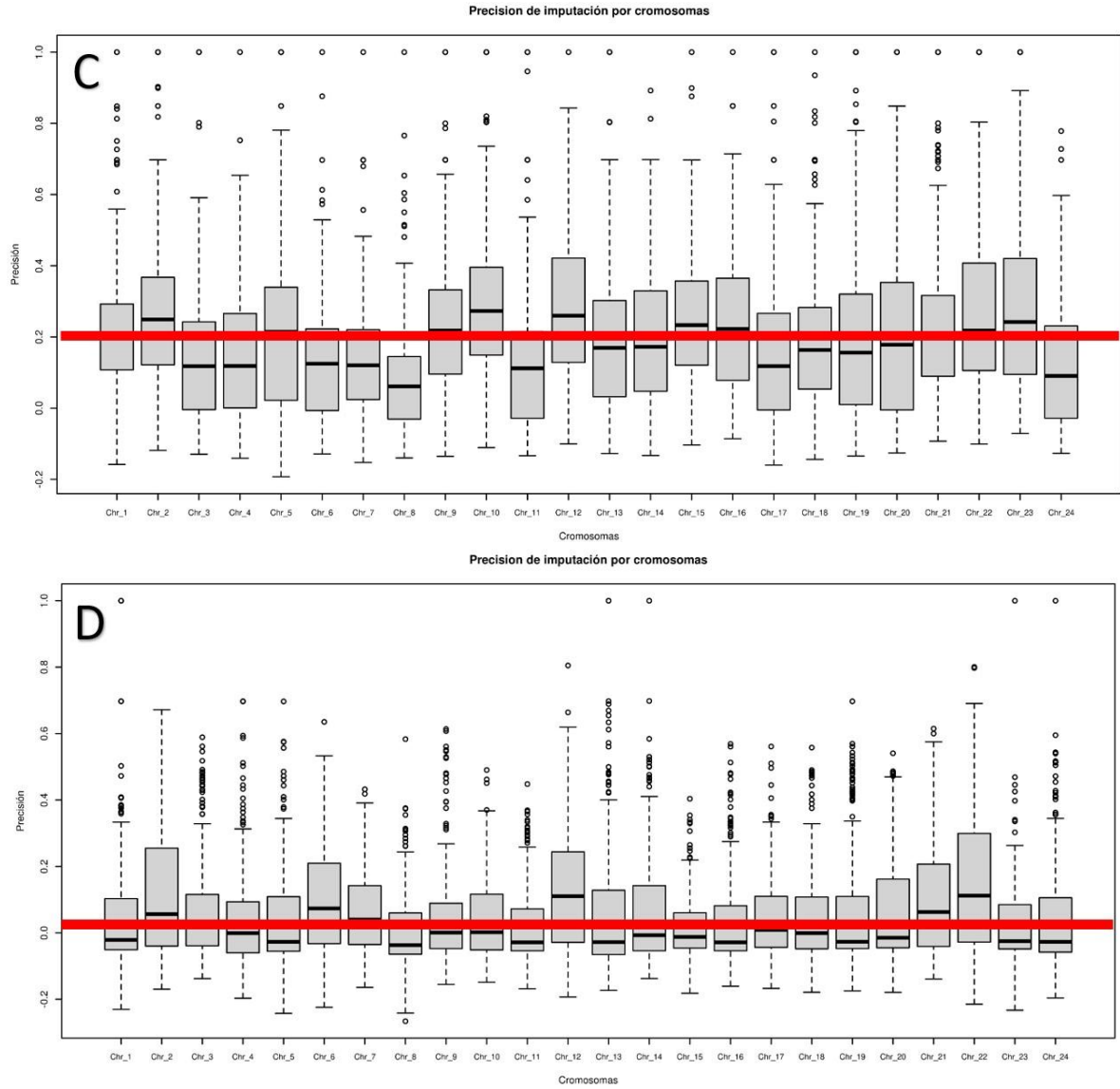


Figura 4: Precisión de imputación según cromosomas para densidades 2K (A), 1K (B), 0,5K (C) y 0,1K (D)

Se estimó el número de marcadores cuya precisión fue mayor al 80% y el porcentaje en función al panel de referencia, según densidad (2K, 1K, 0,5K y 0,1K) (Tabla 5)

Densidad	Marcadores sobre 80%	%
2K	3917	44,04
1K	1865	20,96
0,5K	721	8,11
0,1K	110	1,24

Tabla 5: Número de marcadores con precisión de imputación mayor al 80% según densidad (2K, 1K, 0,5K y 0,1K) y su porcentaje en relación al panel de referencia (9K)

DISCUSIÓN

Se ha reportado que la densidad de los marcadores juega un rol importante al evaluar la precisión de imputación (Sun *et al.*, 2012; Hickey *et al.*, 2012) ya que a mayores densidades de los paneles de baja densidad, mejoran la precisión de imputación, esto se observa en la disminución de la precisión de correlación al reducir de 2K a 0,1K, donde la precisión de imputación disminuye desde 0,53 a 0,06 (Tabla 4 y Figura 4). De la misma manera, como el número de individuos en la población de referencia (individuos genotipados en mayor densidad) impacta la precisión de imputación (Hickey *et al.*, 2012; Sun *et al.*, 2012), los valores actuales de precisión de imputación pueden mejorar al incrementar el número de individuos genotipados usados como referencia, tal como demostraron en su estudio Neuditschko *et al.* (2017).

Al observar la distribución de MAF de los marcadores se observa un amplio número de SNPs que poseen bajo valor de MAF con 3.290 marcadores con un MAF < 0,05 después del control de calidad (Tabla 2). Se ha reportado que marcadores con bajo valor de MAF reducen la precisión de imputación, debido a que la probabilidad de imputar un genotipo homocigoto a otro genotipo homocigoto es mayor para los marcadores con mayor MAF (Ma *et al.*, 2013). El mayor parentesco entre individuos puede mejorar la precisión de imputación (Boison *et al.*, 2015). Sin embargo, al realizar el PCA se observan tres subpoblaciones bien definidas en la población de cobia (Figura 3), lo cual puede haber influido en la precisión de imputación en el presente estudio. Otro factor que también podría mejorar la precisión de imputación es la presencia de información relacionada con el pedigrí de la población (Whalen *et al.*, 2019), como esta información no estaba disponible, esta ausencia disminuyó la precisión de imputación. Finalmente, los marcadores de cobia fueron genotipados con tecnología ddRAD-Seq, esta tecnología permite el genotipado de cantidades masivas de polimorfismos de un solo nucleótido (SNP) para un gran número de individuos y poblaciones a bajo costo, basándose en el uso de doble digestión del genoma, usando dos endonucleasas que actúan en sitios de restricción; posteriormente realiza una selección de los fragmentos digeridos por tamaño y por presencia en sus extremos de ambos cortes de restricción (Peterson *et al.*, 2012). Una de las desventajas de este enfoque es la presencia de regiones genómicas con pocos fragmentos de restricción del tamaño

adecuado para generar marcadores o ausencia de polimorfismos, generando *gaps* en los grupos de ligamiento (Yol *et al.*, 2021), dado que la imputación se basa en el desequilibrio de ligamiento, estos *gaps* pueden afectar la precisión de imputación.

Al analizar la precisión de imputación, el panel de 2K percibe el mayor número de marcadores con una precisión mayor al 80%, pudiendo obtener un panel de 6K aproximadamente (2049 + 3917= 5966) (Tabla 6) a través de la metodología de imputación. Al simular una producción tipo de 225 padres (150 ♀ y 75 ♂) genotipados a mediana densidad (9K) y una progenie de 1000 hijos genotipados tanto a 6K y a 2K con el costo de genotipado según los precios de mercado descritos en la figura 1, se obtiene un costo total de 36.750 USD\$ y 26.750 USD\$ para 6K y 2K de genotipado de la progenie respectivamente, por lo tanto genotipar la progenie con 2K nos permite obtener un panel de 6K con un ahorro del 27% del precio de genotipado total, es decir una reducción de costos debido a la implementación de la imputación.

CONCLUSIONES

- Las precisiones de imputación, de este estudio, si bien no son homologables entre especies y producciones, ya que la imputación depende de la estructura de cada población y el nivel de información genómica disponible, pero los desafíos emergentes para implementar una imputación, como: la densidad de marcadores, tamaño de la muestra, distribución de MAF en los marcadores, parentesco entre individuos, disponibilidad del pedigrí y la tecnología de genotipado, son desafíos probables en otras especies emergentes y producciones. Por lo tanto, los resultados observados en este estudio en cobia, una especie no tradicional y siendo este el primer trabajo basado en el genotipado empleando tecnología ddRAD-seq en cobia, pueden ser usados como referencias para implementar la imputación en próximos proyectos de producciones de especies no tradicionales.
- A partir de los resultados, se puede concluir que genotipar un panel de 2K en la progenie permite la obtención de un panel de SNPs de 6K por menor costo y alta precisión, gracias a la imputación. El presente estudio solo incluyó la precisión de imputación, las diferentes densidades de SNPs y el costo de genotipado para definir la densidad óptima costo-efectiva de genotipado para la progenie, en una producción tipo, sin embargo, a propósito de las evidencias encontradas, el autor sugiere indagar en futuras investigaciones con la inclusión en el presente análisis comparativo, de la precisión de la predicción genómica y como está podría influir en la selección de un panel obtenido mediante imputación.

BIBLIOGRAFÍA

BEUZEN, ND.; STEAR, MJ.; CHANG, KC. 2000. Molecular markers and their use in animal breeding (en línea). s.l., Bailliere Tindall Ltd, vol.160. p. 42-52 DOI: <https://doi.org/10.1053/tvj.2000.0468>.

BOISON, SA.; SANTOS, DJA.; UTSUNOMIYA, AHT.; CARVALHEIRO, R.; NEVES, HHR.; O'BRIEN, AMP.; GARCIA, JF.; SÖLKNER, J.; DA SILVA, MVGB. 2015. Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (*Bos indicus*) dairy cattle: Comparison of commercially available SNP chips. *Journal of Dairy Science* 98(7):4969-4989. DOI: <https://doi.org/10.3168/JDS.2014-9213>.

BOLORMAA, S.; GORE, K.; VAN DER WERF, JHJ.; HAYES, BJ.; DAETWYLER, HD. 2015. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy (en línea). *Animal genetics* 46(5):544-556. DOI: <https://doi.org/10.1111/AGE.12340>.

BROWNING, SR. 2008. Missing data imputation and haplotype phase inference for genome-wide association studies (en línea). s.l., NIH Public Access, vol.124. p. 439-450 DOI: <https://doi.org/10.1007/s00439-008-0568-7>.

CALUS, MPL.; BOUWMAN, AC.; HICKEY, JM.; VEERKAMP, RF.; MULDER, HA. 2014. Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: A review of livestock applications (en línea). *Animal* 8(11):1743-1753. DOI: <https://doi.org/10.1017/S1751731114001803>.

CATCHEN, JM.; AMORES, A.; HOHENLOHE, P.; CRESKO, W.; POSTLETHWAIT, JH. 2011. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences (en línea). *G3 Genes|Genomes|Genetics* 1(3):171-182. DOI: <https://doi.org/10.1534/G3.111.000240>.

CORIALANO, M.C.; COHELO, LCB. 2012. Cobia (*Rachycentron canadum*): a marine fish native to Brazil with biological characteristics to captive environment. (en línea). s.l., s.e. p. 119-132. Consultado 14 sep. 2020. Disponible en https://www.researchgate.net/publication/275354971_Cobia_Rachycentron_canadum_a_marine_fish_native_to_Brazil_with_biological_characteristics_to_captive_environment.

DUFFLOCQ, P.; PÉREZ-ENCISO, M.; LHORENTE, JP.; YÁÑEZ, JM. 2019. Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: A simulation study. *Aquaculture* 503:225-230. DOI: <https://doi.org/10.1016/j.aquaculture.2018.12.061>.

ELY, JJ.; BISHOP, MA.; LAMMEY, ML.; SLEEPER, MM.; STEINER, JM.; LEE, DR. 2010. Use of Biomarkers of Collagen Types I and III Fibrosis Metabolism to Detect Cardiovascular and Renal Disease in Chimpanzees (*Pan troglodytes*) (en línea). *Comparative Medicine* 60(2):154. Consultado 14 jul. 2022. Disponible en </pmc/articles/PMC2855044/>.

FAO. 2018. El estado mundial de la pesca y la acuicultura 2018. Cumplir los objetivos de

desarrollo sostenible. s.l., s.e. 250 p. DOI: <https://doi.org/CC BY-NC-SA 3.0 IGO>.

FAO. 2020. Fisheries Global Information System (FAO-FIGIS) (en línea, sitio web). Consultado 14 sep. 2020. Disponible en <http://www.fao.org/fishery/>.

GEORGES, M.; CHARLIER, C.; HAYES, B. 2018. Harnessing genomic information for livestock improvement (en línea). *Nature Reviews Genetics* 2018 20:3 20(3):135-156. DOI: <https://doi.org/10.1038/s41576-018-0082-2>.

GJEDREM, T. 2012. Genetic improvement for the development of efficient global aquaculture: A personal opinion review. s.l., Elsevier, vols.344-349. p. 12-22 DOI: <https://doi.org/10.1016/j.aquaculture.2012.03.003>.

GODDARD, ME.; HAYES, BJ. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes (en línea). s.l., Nature Publishing Group, vol.10. p. 381-391 DOI: <https://doi.org/10.1038/nrg2575>.

GUTIÉRREZ, MSERF. 2017. LA COBIA (*Rachycentron canadum*) (en línea). *Revista Cubana de Investigaciones Pesqueras*, :0138-8452. Consultado 14 sep. 2020. Disponible en https://scholar.googleusercontent.com/scholar?q=cache:eYEURbiWHJkJ:scholar.google.com/+cobia+produccion&hl=es&as_sdt=0,5.

HALDAR, C. 2018. Importance of Marker Assisted Selection in Fish Breeding in India: Challenges and Opportunities (en línea). *Annals of Aquaculture and Research* 5((1)):1047. Consultado 8 jun. 2022. Disponible en <https://agris.fao.org/agris-search/search.do?recordID=US2019V00068>.

HARVEY, BJ.; SOTO, D.; CAROLSFELD, J.; BEVERIDGE, MCM.; BARTLEY, DM. 2016. Planning for aquaculture diversification : the importance of climate change and other drivers : FAO Technical Workshop, 23-25 June 2016, FAO Rome, Italy. s.l., s.e. 154 p.

HAYES, B.; GODDARD, M. 2010. Genome-wide association and genomic selection in animal breeding (en línea). *In Genome*. s.l., Genome. p. 876-883 DOI: <https://doi.org/10.1139/G10-076>.

HAYES, BJ.; BOWMAN, PJ.; DAETWYLER, HD.; KIJAS, JW.; VAN DER WERF, JHJ. 2012. Accuracy of genotype imputation in sheep breeds (en línea). *Animal Genetics* 43(1):72-80. DOI: <https://doi.org/10.1111/j.1365-2052.2011.02208.x>.

HEIDARITABAR, M.; CALUS, MPL.; VEREIJKEN, A.; GROENEN, MAM.; BASTIAANSEN, JWM. 2015. Accuracy of imputation using the most common sires as reference population in layer chickens (en línea). *BMC Genetics* 16(1):101. DOI: <https://doi.org/10.1186/s12863-015-0253-5>.

HENDERSON, CR. 1984. Applications of linear models in animal breeding. Applications of linear models in animal breeding. .

HICKEY, JM.; CROSSA, J.; BABU, R.; DE LOS CAMPOS, G. 2012. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs.

Crop Science 52(2):654-663. DOI: <https://doi.org/10.2135/CROPSCI2011.07.0358>.

HILL, WG.; ROBERTSON, A. 1968. Linkage disequilibrium in finite populations (en línea). *Theoretical and Applied Genetics* 1968 38:6 38(6):226-231. DOI: <https://doi.org/10.1007/BF01245622>.

HOZÉ, C.; FOUILLOUX, MN.; VENOT, E.; GUILLAUME, F.; DASSONNEVILLE, R.; FRITZ, S.; DUCROCQ, V.; PHOCAS, F.; BOICHARD, D.; CROISEAU, P. 2013. High-density marker imputation accuracy in sixteen French cattle breeds (en línea). *Genetics Selection Evolution* 45(1):33. DOI: <https://doi.org/10.1186/1297-9686-45-33>.

HUANG, L.; LI, Y.; SINGLETON, AB.; HARDY, JA.; ABECASIS, G.; ROSENBERG, NA.; SCHEET, P. 2009. Genotype-Imputation Accuracy across Worldwide Human Populations. *The American Journal of Human Genetics* 84(2):235-250. DOI: <https://doi.org/10.1016/J.AJHG.2009.01.013>.

HUANG, Y.; HICKEY, JM.; CLEVELAND, MA.; MALTECCA, C. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost (en línea). *Genetics Selection Evolution* 44(1). DOI: <https://doi.org/10.1186/1297-9686-44-25>.

JAFARI, M.; ANSARI-POUR, N. 2019. Why, When and How to Adjust Your P Values? (en línea). *Cell Journal (Yakhteh)* 20(4):604. DOI: <https://doi.org/10.22074/CELLJ.2019.5992>.

JOAN HOLT, G.; FAULK, CK.; SCHWARZ, MH. 2007. A review of the larviculture of cobia *Rachycentron canadum*, a warm water marine fish. s.l., Elsevier, vol.268. p. 181-187 DOI: <https://doi.org/10.1016/j.aquaculture.2007.04.039>.

KOOPAEE, HK.; KOSHKOIYEH, AE. 2014. SNPs genotyping technologies and their applications in farm animals breeding Programs: Review (en línea). *Brazilian Archives of Biology and Technology* 57(1):87-95. DOI: <https://doi.org/10.1590/S1516-89132014000100013>.

LACHANCE, J. 2016. Hardy-Weinberg Principle - an overview | ScienceDirect Topics (en línea). s.l., s.e. Consultado 21 sep. 2020. Disponible en <https://www.sciencedirect.com/topics/neuroscience/hardy-weinberg-principle>.

LANGMEAD, B.; SALZBERG, SL. 2012. Fast gapped-read alignment with Bowtie 2 (en línea). *Nature Methods* 2012 9:4 9(4):357-359. DOI: <https://doi.org/10.1038/nmeth.1923>.

MA, P.; BRØNDUM, RF.; ZHANG, Q.; LUND, MS.; SU, G. 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *Journal of Dairy Science* 96(7):4666-4677. DOI: <https://doi.org/10.3168/JDS.2012-6316>.

MAREES, AT.; DE KLUIVER, H.; STRINGER, S.; VORSPAN, F.; CURIS, E.; MARIE-CLAIRE, C.; DERKS, EM. 2018. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis (en línea). *International Journal of Methods in Psychiatric Research* 27(2). DOI: <https://doi.org/10.1002/mpr.1608>.

MEUWISSEN, THE.; HAYES, BJ.; GODDARD, ME. 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157(4).

MUNDO-ACUÍCOLA. 2015. Diversificación Acuícola: La experiencia que dejó el cultivo de cobia en Mejillones. (111):44.

NEUDITSCHKO, M.; RAADSMA, HW.; KHATKAR, MS.; JONAS, E.; STEINIG, EJ.; FLURY, C.; SIGNER-HASLER, H.; FRISCHKNECHT, M.; VON NIEDERHÄUSERN, R.; LEEB, T.; RIEDER, S. 2017. Identification of key contributors in complex population structures (en línea). *PLOS ONE* 12(5):e0177638. DOI: <https://doi.org/10.1371/JOURNAL.PONE.0177638>.

OYARZÚN, PA.; TORO, JE.; NAVARRO, JM. 2013. Comparison of the physiological energetics between *Mytilus chilensis*, *Mytilus galloprovincialis* and their hybrids, under laboratory conditions. *Aquaculture Research* 44(12):1805-1814. DOI: <https://doi.org/10.1111/j.1365-2109.2012.03184.x>.

PEREZ-ENRIQUEZ, R.; MAX-AGUILAR, A. 2016. Rastreabilidad del pedigrí en camarón blanco (*Litopenaeus vannamei*) mediante marcadores genéticos: Una comparación entre microsatélites y SNP (en línea). *Ciencias Marinas* 42(4):227-235. DOI: <https://doi.org/10.7773/cm.v42i4.2662>.

PETERSON, BK.; WEBER, JN.; KAY, EH.; FISHER, HS.; HOEKSTRA, HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species (en línea). *PloS one* 7(5). DOI: <https://doi.org/10.1371/JOURNAL.PONE.0037135>.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, MAR.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, PIW.; DALY, MJ.; SHAM, PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses (en línea). *American Journal of Human Genetics* 81(3):559-575. DOI: <https://doi.org/10.1086/519795>.

R CORE TEAM. 2022. R: The R Project for Statistical Computing (en línea, sitio web). Consultado 4 jul. 2022. Disponible en <https://www.r-project.org/>.

RSTUDIO. 2020. RStudio | Open source & professional software for data science teams - RStudio (en línea, sitio web). Consultado 11 nov. 2020. Disponible en <https://rstudio.com/>.

SARGOLZAEI, M.; CHESNAIS, JP.; SCHENKEL, FS. 2014. A new approach for efficient genotype imputation using information from relatives (en línea). *BMC Genomics* 15(1):478. DOI: <https://doi.org/10.1186/1471-2164-15-478>.

SHAFFER, RV.; NAKAMURA, EL. 1989. Synopsis of biological data on the cobia *Rachycentron canadum* (Pisces: Rachycentridae). .

SLATKIN, M. 2008. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future (en línea). *Nature Reviews Genetics* 2008 9:6 9(6):477-485. DOI: <https://doi.org/10.1038/nrg2361>.

SPENCER, CCA.; SU, Z.; DONNELLY, P.; MARCHINI, J. 2009. Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip (en línea). *PLoS Genetics* 5(5):e1000477. DOI: <https://doi.org/10.1371/journal.pgen.1000477>.

SUBPESCA. 2020. Informe sectorial de pesca y acuicultura. Subsecretaría de Pesca y Acuicultura. Departamento de Análisis Sectorial.Chile. :20.

SUN, C.; WU, XL.; WEIGEL, KA.; ROSA, GJM.; BAUCK, S.; WOODWARD, BW.; SCHNABEL, RD.; TAYLOR, JF.; GIANOLA, D. 2012. An ensemble-based approach to imputation of moderate-density genotypes for genomic selection with application to Angus cattle (en línea). *Genetics research* 94(3):133-150. DOI: <https://doi.org/10.1017/S001667231200033X>.

TSAI, HY.; MATIKA, O.; EDWARDS, SMK.; ANTOLÍN-SÁNCHEZ, R.; HAMILTON, A.; GUY, DR.; TINCH, AE.; GHARBI, K.; STEAR, MJ.; TAGGART, JB.; BRON, JE.; HICKEY, JM.; HOUSTON, RD. 2017. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon (en línea). *G3: Genes, Genomes, Genetics* 7(4):1377-1383. DOI: <https://doi.org/10.1534/g3.117.040717>.

UEMOTO, Y.; SASAKI, S.; SUGIMOTO, Y.; WATANABE, T. 2015. Accuracy of high-density genotype imputation in Japanese Black cattle (en línea). *Animal Genetics* 46(4):388-394. DOI: <https://doi.org/10.1111/age.12314>.

VALLEJO, RL.; LEEDS, TD.; GAO, G.; PARSONS, JE.; MARTIN, KE.; EVENHUIS, JP.; FRAGOMENI, BO.; WIENS, GD.; PALTÍ, Y. 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture (en línea). *Genetics Selection Evolution* 49(1):1-13. DOI: <https://doi.org/10.1186/s12711-017-0293-6>.

VENTURA, R V.; LU, D.; SCHENKEL, FS.; WANG, Z.; LI, C.; MILLER, SP. 2014. Impact of reference population on accuracy of imputation from 6K to 50K single nucleotide polymorphism chips in purebred and crossbreed beef cattle (en línea). *Journal of animal science* 92(4):1433-1444. DOI: <https://doi.org/10.2527/JAS.2013-6638>.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics (en línea). *Genetics Selection Evolution* 34(3):275. DOI: <https://doi.org/10.1186/1297-9686-34-3-275>.

WANG, Y.; CAI, Z.; STOTHARD, P.; MOORE, S.; GOEBEL, R.; WANG, L.; LIN, G. 2012. Fast accurate missing SNP genotype local imputation (en línea). *BMC Research Notes* 5(1):404. DOI: <https://doi.org/10.1186/1756-0500-5-404>.

WANG, Y.; WYLIE, T.; STOTHARD, P.; LIN, G. 2015. Whole genome SNP genotype piecemeal imputation (en línea). *BMC Bioinformatics* 16(1):1-11. DOI: <https://doi.org/10.1186/s12859-015-0770-2>.

WENNE, R. 2018. Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations (en línea). s.l., EDP

Sciences, vol.31. p. 2 DOI: <https://doi.org/10.1051/alr/2017043>.

WHALEN, A.; GORJANC, G.; HICKEY, JM. 2019. Family-specific genotype arrays increase the accuracy of pedigree-based imputation at very low marker densities (en línea). *Genetics Selection Evolution* 51(1):1-9. DOI: <https://doi.org/10.1186/S12711-019-0478-2/FIGURES/4>.

YÁÑEZ, JM.; NEWMAN, S.; HOUSTON, RD. 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. s.l., Frontiers Research Foundation, vol.6. DOI: <https://doi.org/10.3389/fgene.2015.00128>.

YANG, W.; YANG, Y.; ZHAO, C.; YANG, K.; WANG, D.; YANG, J.; NIU, X.; GONG, J. 2020. Animal-ImputeDB: A comprehensive database with multiple animal reference panels for genotype imputation (en línea). *Nucleic Acids Research* 48(D1):D659-D667. DOI: <https://doi.org/10.1093/nar/gkz854>.

YOL, E.; BASAK, M.; KIZIL, S.; LUCAS, SJ.; UZUN, B. 2021. A High-Density SNP Genetic Map Construction Using ddRAD-Seq and Mapping of Capsule Shattering Trait in Sesame. *Frontiers in Plant Science* 12:917. DOI: <https://doi.org/10.3389/FPLS.2021.679659/BIBTEX>.

YOSHIDA, GM.; LHORENTE, JP.; CORREA, K.; SOTO, J.; SALAS, D.; YÁÑEZ, JM. 2019. Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*) (en línea). *G3: Genes, Genomes, Genetics* 9(8):2597-2607. DOI: <https://doi.org/10.1534/g3.119.400116>.

YOU, X.; SHAN, X.; SHI, Q. 2020. Research advances in the genomics and applications for molecular breeding of aquaculture animals. s.l., Elsevier B.V., vol.526. p. 735357 DOI: <https://doi.org/10.1016/j.aquaculture.2020.735357>.