



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS VETERINARIAS y PECUARIAS
ESCUELA DE POSTGRADO Y POSTÍTULO

PRECISIÓN EN LA PREDICCIÓN GENÓMICA UTILIZANDO
DIFERENTES NIVELES DE ERROR DE IMPUTACIÓN EN
PROGRAMAS DE MEJORAMIENTO GENÉTICO
ACUÍCOLAS

PABLO RAFAEL DUFFLOCQ URMENETA

Tesis para optar al Grado de
Magíster en Ciencias Animales y Veterinarias y al
Título Profesional de Médico Veterinario

Santiago-Chile

2017



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS VETERINARIAS y PECUARIAS
ESCUELA DE POSTGRADO Y POSTÍTULO

**PRECISIÓN EN LA PREDICCIÓN GENÓMICA UTILIZANDO
DIFERENTES NIVELES DE ERROR DE IMPUTACIÓN EN
PROGRAMAS DE MEJORAMIENTO GENÉTICO
ACUÍCOLAS**

PABLO RAFAEL DUFFLOCQ URMENETA

Tesis para optar al Grado de
Magíster en Ciencias Animales y Veterinarias y al
Título Profesional de Médico Veterinario

DIRECTOR DE TESIS: DR. JOSÉ MANUEL YÁÑEZ L.

Santiago-Chile

2017



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS VETERINARIAS y PECUARIAS
ESCUELA DE POSTGRADO Y POSTÍTULO**

INFORME DE APROBACIÓN DE TESIS DE MAGÍSTER

SE INFORMA A LA DIRECCIÓN DE POSTGRADO Y POSTÍTULO DE LA FACULTAD DE CIENCIAS VETERINARIAS Y PECUARIAS, QUE LA TESIS DE MAGÍSTER PRESENTADA POR EL CANDIDATO

PABLO RAFAEL DUFFLOCQ URMENETA

HA SIDO APROBADA POR LA COMISIÓN EVALUADORA DE TESIS COMO REQUISITO PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS ANIMALES Y VETERINARIAS EN EXAMEN DE DEFENSA DE TESIS RENDIDO EL DÍA 04 DE ENERO DE 2018

DIRECTOR DE TESIS

DR. JOSÉ MANUEL YÁÑEZ L.

COMISIÓN EVALUADORA E INFORMANTE DE TESIS

DR RICARDO VERDUGO S.

DR. JURIJ WACYK G.

Esta Tesis de Grado se realizó en el Laboratorio de Genética y Genómica Acuícola, del Departamento de Medicina Preventiva Animal de la Facultad de Ciencias Veterinarias y Pecuarias de la Universidad de Chile, y contó con financiamiento

PROYECTO FONDEF IT/10100

DEDICATORIA

Esta memoria de título está dedicada a mi familia y amigos, por ser un apoyo fundamental en esta constante búsqueda de conocimiento y de persecución de metas en el camino que elegí de la medicina veterinaria.

Pablo Rafael Dufflocq Urmeneta

AGRADECIMIENTOS

Me gustaría agradecer a la Facultad de Ciencias Veterinarias y Pecuarias de la Universidad de Chile por entregarme las herramientas y formación necesaria para desenvolverme como futuro profesional íntegro y consiente de mi entorno tanto social como físico, con el objetivo de ser un profesional que aporte al desarrollo científico y tecnológico nacional e internacional.

Agradezco al Fondo de Fomento al Desarrollo Científico y Tecnológico (FONDEF) por el financiamiento del proyecto de investigación que me permitió realizar la presente memoria de título y a las empresas Aguas Claras S.A. y AquaInnovo S.A. por facilitar tanto los peces como los datos utilizados para los análisis posteriores.

Agradezco al programa de Becas Santander por la oportunidad brindada para realizar pasantía de estudio en el Centre de Recerca Agrigenòmica (CRAG) de la Universidad Autònoma de Barcelona, donde pude realizar parte de mi tesis y vivir una experiencia que me ayudó a crecer en múltiples aspectos de manera inmensamente positiva.

De forma especial quiero agradecer al Dr. José Manuel Yáñez por su apoyo incondicional y consejo desde el pregrado hasta la fecha, por darme la confianza y la oportunidad de formar parte de su equipo de trabajo en el laboratorio de Genética y Genómica Acuícola y de desarrollarme en las áreas de genética y estadística. Muchas gracias también por el seguimiento, consejos y compromiso con esta memoria de título, en la cual hay mucho tiempo de formación, aprendizaje y esfuerzo previo, la cual concluyó de manera satisfactoria. Muchas gracias por la amistad y el apoyo constante que me has entregado.

También agradezco a los miembros de la comisión evaluadora de esta memoria, al Dr. Jurij Wacyk y Dr. Ricardo Verdugo por ser parte importante de este proceso, los cuales siempre velaron por la correcta ejecución de esta memoria desde sus consejos y guía académica.

De manera especial me gustaría agradecer al Dr. Miguel Pérez-Enciso, investigador del laboratorio CRAG, el cual fue mi tutor durante mi estancia en Barcelona. Muchas gracias doc por la amistad, guía constante, buena disposición para ayudarme y responder cualquier duda en relación a los temas abordados antes, durante y después de mi estancias por sus

tierras. También me gustaría hacer una mención especial al Dr. Sebastián Ramos-Onsins, Dra. Sara Guirao, Dr. José María Folch por su cálida acogida, confianza y simpatía desde que llegué al laboratorio y a Barcelona. También quiero agradecer a mi amigo el Dr. Lifei Ma por su apoyo y simpatía. De manera muy especial me gustaría agradecer a la Dra. Natalia Forneris quién fue una excelente guía académica y compañera, por su apoyo constante e incondicional el cual recuerdo con mucho cariño.

Quiero agradecer a Ana “Moni” Jedlicki Corbeaux su apoyo incondicional, cariño y confianza que me ha entregado siempre, pues ella ha sido un apoyo inmenso en el laboratorio del cual formo parte. Muchas gracias por ser la maravillosa persona que conozco y ser un pilar fundamental en el laboratorio.

Me gustaría agradecer enormemente a mis compañeros de laboratorio el Dr. Agustín “Shasho” Barría, a quien he tenido el agrado de conocer y compartir desde hace ya algunos años, gracias por toda la buena onda, consejos y apoyo constante. A la Dra. María Eugenia López, por su simpatía y buena disposición para ayudarme a resolver dudas que aparecieron durante la realización de esta memoria. A mis compañeros Halley, Raúl, Lianne, Catalina, Khatarina, Grazyella, Rafael, Dicrán, Marzelo, Giovanna, Isadora, María Ignacia con quienes he pasado buenos momentos de trabajo y de amistad.

Quisiera agradecer de manera especial al Dr. Luis Ibarra Martínez, quién me incentivó y apoyó en instancias iniciales de mi paso por el pregrado para desarrollarme en el área de la estadística aplicada a la medicina veterinaria, herramienta que me ha abierto varias puertas y brindado oportunidades de desarrollarme en esa área desde la academia que quizás de otra manera hubiese sido imposible lograr.

También no puedo dejar de agradecer a mis amigos de Favet, Felipe Cerpa, Emilio Aguilar y Felipe Pizarro (“Los Piripis”), a Javiera Fuentes y Gabriela Boldt quienes siempre me han apoyado y dado ánimos para continuar y alcanzar las metas que me he propuesto. Muchas gracias por ser un pilar importante en este proceso y durante todo mi paso por la universidad. Quiero mencionar a mis amigos de Construcción Civil de la Pontificia Universidad Católica, con quienes mantengo una gran amistad hasta el día de hoy, muchas gracias por el apoyo

constante Marco, Kevin, Ariel, Catita, Álvaro, Vale, Edu, Jenni, Gonza (y a sus lindas hijitas) y por todo el cariño que me han demostrado a pesar de haber elegido un camino diferente al que comenzamos juntos. También me gustaría agradecer a mis amigos de la Academia de Kung-Fu Shen-Men Maipú, en especial al sifu José Luis Marchant, por su apoyo incondicional y siempre impulsarnos a la superación constante.

Finalmente quiero agradecer infinitamente a mis padres Regina y Julio, mis hermanos Luis, Julio y María Luisa, por todo el apoyo que me han brindado desde siempre tanto desde la exigencia como del cariño que percibo desde ellos, a mis sobrinos Catalina, Francisca, Julio, Mateo, Maite, Raimundo, Tomás y Martina, cuñadas Angélica y Marcela, tía Lole y primos Daniela y Gabriel por su apoyo incondicional para poder alcanzar las metas que me he propuesto. Muchas gracias a todos.

Pablo Rafael Dufflocq Urmeneta

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
2. REVISIÓN BIBLIOGRÁFICA	3
2.1 Situación mundial y nacional de la acuicultura.....	3
2.2 Mejoramiento genético en salmónidos.....	4
2.3 Resistencia a enfermedades en peces.....	5
2.4 Marcadores moleculares.....	6
2.5 Selección genómica.....	8
2.6 Imputación de genotipos.....	9
2.7 Estudios de simulación.....	10
3. HIPÓTESIS	12
4. OBJETIVOS	12
4.1 Objetivo General.....	12
4.2 Objetivos Específicos.....	12
5. MATERIAL Y MÉTODOS	13
5.1 Software.....	13
5.2 Población.....	13
5.3 Genotipado.....	13
5.4 Simulación.....	14
5.5 Escenarios.....	15
5.6 Análisis estadístico.....	16
5.6.1 Estimación de valores genéticos y fenotípicos.....	16
5.6.2 Estimación de Precisión.....	18
5.6.3 Estimación de Exactitud.....	19
5.6.4 Evaluación Económica.....	19
6. RESULTADOS	21
6.1 BLUP versus GBLUP42K.....	21
6.2 Efecto del error de imputación.....	23
6.3 Evaluación Económica.....	27
7. DISCUSIÓN	29
8. CONCLUSIONES	36
9. BIBLIOGRAFÍA	37

ÍNDICE DE TABLAS

TABLA 1. Análisis de sesgo para EBVs en la primera (F1) y quinta generación (F5) considerando valores de heredabilidad de 0.1 (baja), 0.2 (media) y 0.4 (alta) por medio de métodos BLUP y GBLUP.....24

TABLA 2. Evaluación económica de implementación de selección genómica utilizando cuatro paneles de SNPs de distinta densidad (0.5K, 3K, 7K y 42K) en comparación con evaluación genética basada en información fenotípica durante un año productivo normal.....28

ÍNDICE DE AYUDAS ILUSTRATIVAS

FIGURA 1. Valores de precisión de EBVs en evaluación genómica considerando diferentes densidades de marcadores (GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K) a través de cinco generaciones, tres niveles de heredabilidad A ($h^2 = 0.1$), B ($h^2 = 0.2$) y C ($h^2 = 0.4$), comparadas con valores de precisión obtenidos por medio del método BLUP.....22

FIGURA 2.1. Representación gráfica de dispersión de TBVs versus EBVs en la primera generación (F1) considerando valores de heredabilidad de 0.1 (baja), 0.2 (media) y 0.4 (alta) por medio de métodos BLUP y GBLUP, con sus respectivas rectas de regresión lineal real (color negro) y recta de regresión ideal ($\beta=1$, color rojo).....26

FIGURA 2.2. Representación gráfica de dispersión de TBVs versus EBVs en la quinta generación (F5) considerando valores de heredabilidad de 0.1 (baja), 0.2 (media) y 0.4 (alta) por medio de métodos BLUP y GBLUP, con sus respectivas rectas de regresión lineal real (color negro) y recta de regresión ideal ($\beta=1$, color rojo).....27

ÍNDICE DE ABREVIATURAS

ADN	: Ácido Desoxirribonucleico
ANOVA	: <i>Analysis of Variance</i>
β	: Coeficiente de regresión lineal
BLUP	: <i>Best Linear Unbiased Prediction</i>
EBV	: <i>Estimated Breeding Value</i>
F	: Generación
GBLUP	: <i>Genomic Best Linear Unbiased Prediction</i>
GWAS	: <i>Genome-Wide Association Studies</i>
h^2	: Heredabilidad
IBD	: <i>Identity By Descent</i>
IPN	: Necrosis Pancreática Infecciosa
£	: Libra Esterlina
LD	: <i>Linkage Disequilibrium</i>
LSD:	: <i>Least Significant Difference</i>
Mt	: Millones de toneladas
NGS	: <i>Next Generation Sequencing</i>
p	: <i>p-value</i>
QTL	: <i>Quantitative Trait Loci</i>
SNP	: <i>Single Nucleotide Polymorphism</i>
SRS	: Síndrome Rickettsial del Salmón
TBV	: <i>True Breeding Value</i>
USD	: <i>United States Dollar (\$)</i>
VIF	: <i>Variance Inflation Factor</i>

RESUMEN

Para caracteres de resistencia a enfermedades dentro del objetivo de mejoramiento en programas de selección genética acuícolas, los candidatos a selección son evaluados usando información fenotípica de sus parientes para obtener los valores de cría (EBVs) por medio del método BLUP. La obtención de fenotipos implica el sacrificio de los peces antes de que se puedan reproducir, por lo que el uso de información molecular para su evaluación genética se está volviendo una práctica común en especies acuícolas de interés comercial. La selección genómica (GBLUP) es un método que usa información genotípica desde paneles de SNPs, la que podría aumentar la precisión con que se seleccionan individuos. Una limitante de esto, es el elevado costo en la obtención de esta información, pero la imputación de genotipos puede representar una solución a bajo costo para resolver este problema. La presente tesis permitió evaluar y comparar, por medio de la simulación de evaluación genómica, los valores de precisión para EBVs usando los métodos BLUP y GBLUP como una aproximación inicial para estudiar la inclusión de dicha información. Se simularon 5 generaciones de 3,600 individuos de trucha arcoíris (*Oncorhynchus mykiss*) cada una desde una población fundadora de 831 padres y 831 madres los cuales cuentan con datos genotípicos reales. Las posiciones de los QTLs se muestrearon aleatoriamente en cada cromosoma (4 QTLs por cromosoma, 116 QTLs totales) desde 42,822 SNP disponibles. Los escenarios variaron de acuerdo con tres niveles de heredabilidad (0.1, 0.2 y 0.4), cuatro densidades de SNPs (0.5K, 3K, 7K y 42K), con un error de imputación asociado a cada uno (10%, 5%, 1% y 0%, respectivamente). Las simulaciones, análisis estadístico posterior y evaluación económica mostraron que: (1) Existen diferencias significativas ($p < 0.05$) al comparar valores de precisión de BLUP y GBLUP en todos los escenarios; (2) un incremento no-lineal en los valores de precisión de GBLUP a través de las diferentes densidades de SNPs y niveles de heredabilidad; (3) se detectaron diferencias significativas entre precisión y exactitud de EBVs al comparar BLUP y GBLUP, en todos los casos; (4) no se detectaron diferencias significativas entre precisión de paneles 0.5K, 3K y 7K SNPs, especialmente a valores de h^2 medias y altas. (5) Debido a esto, sería suficiente contar con 500 SNPs para realizar selección genómica de manera rentable y con mayor precisión que utilizando el método BLUP.

Palabras Claves: selección genómica; simulación; precisión; heredabilidad; densidad de panel de SNP; error de imputación; programa de selección acuícola.

ABSTRACT

For disease resistance traits included in aquaculture breeding programs, candidates for selection are tested using phenotypic information of relatives to obtain breeding values by the BLUP method because it implies the slaughtering of fishes before reproduction. Therefore, the use of molecular information for genetic evaluations is becoming a common practice in most relevant aquaculture species. Genomic selection (GBLUP) is a methodology that uses genotypic information obtained from single nucleotide polymorphism (SNP) panels and it could increase the accuracy of individuals' selection. A limitation of this method is the high costs of obtaining molecular information, but the imputation of genotypes could be a low-cost solution. In the present study was evaluated and compared the precision and accuracy of EBVs using BLUP and GBLUP methods at varying levels of genotypes imputation as an approach to study the inclusion of this molecular information. We simulated 5 generations of 3,600 rainbow trout (*Oncorhynchus mykiss*) each one from 831 sires and 831 dams as founder population with real genotypic data from. The QTLs positions were randomly sampled in each chromosome (4 QTLs per chromosome, 116 QTLs in total) from 42,822 available SNPs. The scenarios vary in heritability levels (0.1, 0.2 and 0.4) and density of SNPs panel (0.5K, 3K, 7K and 42K), rendering varying error rates (10%, 5%, 1% and 0%, respectively). The simulations, statistical analyses and economic evaluation showed: (1) significant differences ($p < 0.05$) in precision between BLUP and GBLUP estimates in all scenarios; (2) a non-linear increase in GBLUP precision with SNP density and heritability levels; (3) significant differences were found between BLUP and GBLUP in precision and accuracy of EBVs, (4) non-significant differences between GBLUP0.5K, GBLUP3K and GBLUP7K precisions, especially for medium and high heritability levels. (5) Consequently, a 500 SNPs panel would be enough to obtain higher precision of EBVs from genomic selection than from pedigree-based method in a rentable way.

Keywords: genomic selection; simulation; precision; heritability; SNP panel density; imputation error; aquaculture breeding programs.

INTRODUCCIÓN

En esquemas de selección acuícolas, algunas características de interés comercial son evaluadas en hermanos completos o medios hermanos de los candidatos y esta información se usa para calcular los valores de cría con los que se realizará la selección de los reproductores (Gjedrem, 2012). Por ejemplo, los caracteres de resistencia a enfermedades son imposibles de evaluar en los candidatos a la selección debido a que la obtención de registros para estas características implican el sacrificio de los animales evaluados (Nielsen *et al.*, 2009). Para estas características, la predicción del mérito genético se basa en información proveniente del pedigrí e información fenotípica de los hermanos completos y medios hermanos de los candidatos a selección por medio del método BLUP (*Best Linear Unbiased Prediction*) (Henderson, 1984). En este caso, la mitad de la variación genética (entre familias) es explotada (Sonesson y Meuwissen, 2009). Debido a la creciente disponibilidad de tecnologías de genotipado, el uso de información molecular para evaluación genética se está convirtiendo en la metodología estándar en la mayoría de las especies de interés comercial (Hayes *et al.*, 2009). En programas de selección acuícola, la selección genómica puede aumentar la precisión con que se seleccionan los individuos para características de relevancia comercial utilizando la variación genética entre y dentro de familias, con el respectivo incremento en el progreso genético (Sonesson y Meuwissen, 2009). La selección genómica explota información desde los marcadores moleculares denominados SNP (*Single Nucleotide Polymorphism*) y la información fenotípica para predecir el mérito genético de un animal (Meuwissen *et al.*, 2001). Varios paneles de SNPs han sido desarrollados para especies de interés comercial tales como bovinos (Matukumalli *et al.*, 2009), cerdos (Ramos *et al.*, 2009), aves (Groenen *et al.*, 2011), salmón del Atlántico (Houston *et al.*, 2014; Yáñez *et al.*, 2014c; Yáñez *et al.*, 2016) y trucha arcoíris (Palti *et al.*, 2015). Sin embargo, la limitante asociada a los costos de adquisición o desarrollo de éstos, es un aspecto relevante a considerar (Huang *et al.*, 2012).

El uso de paneles de SNP de baja densidad en conjunto con estrategias de imputación de genotipos, representa una alternativa que podría disminuir los costos de genotipado para evaluación genómica en especies terrestres (Cleveland y Hickey, 2013) y acuícolas (Tsai *et al.*, 2017; Yoshida *et al.*, submitted). Varios algoritmos han sido desarrollados para imputar

genotipos, tales como Beagle (Browning y Browning, 2007), AlphaImpute (Hickey *et al.*, 2012) y FImpute (Sargolzaei *et al.*, 2014), los que difieren en la precisión y rapidez con que realizan el computo (Huang *et al.*, 2012). Por su parte, la precisión con que éstos se imputan, va a depender de la estrategia de genotipado que se use, el uso de paneles de SNP de alta, media o baja densidad, el número de individuos que se utilice como referencia y el parentesco de los individuos en la población de referencia y validación (Huang *et al.*, 2012; Cleveland and Hickey, 2013; Tsai *et al.*, 2017; Yoshida *et al.*, submitted).

El uso de simulaciones es una manera común, a bajo costo y rápida para investigar la óptima incorporación de información molecular en esquemas de selección animal (Nordborg y Tavaré, 2002). Diferentes escenarios se pueden evaluar para caracteres de interés comercial considerando diferente número de generaciones, heredabilidad, paneles de SNP de distinta densidad, errores de imputación, información familiar, arquitectura genética, desequilibrio de ligamiento y forma en que los SNPs se distribuyen a lo largo del genoma. Para esto, existen varios programas de simulación como AlphaMPSim (Hickey *et al.*, 2014), XSim (Cheng *et al.*, 2015) y QMSim (Sargolzaei y Schenkel, 2009), los que requieren de una exploración previa de los parámetros usados en las simulaciones para generar escenarios lo más realistas posible en especies con relativa ausencia de conocimiento e incertidumbre sobre su historia demográfica reciente y pasada (por ejemplo, cuellos de botella, cruzamientos o selección intensiva), como es el caso de la mayoría de las especies acuícolas. Para resolver esto, el programa “ms2gs” fue desarrollado como una herramienta flexible, la cual utiliza el método de “*gene-dropping*” para simular caracteres complejos y realizar su evaluación genómica en diferentes escenarios utilizando los niveles de diversidad genética y LD (*Linkage Disequilibrium*) reales de la población mediante el uso de genotipos reales como base para la simulación (Pérez-Enciso y Legarra, 2016).

El objetivo de este estudio fue determinar la precisión de la selección genómica al estimar valores de cría y compararla con valores de precisión obtenidos por el método BLUP en distintos escenarios, los que consideran la simulación de paneles de SNPs de diferente densidad a los que se les asoció un nivel de error de imputación específico, a lo largo de distinto número de generaciones y valores de heredabilidad.

REVISIÓN BIBLIOGRÁFICA

Situación mundial y nacional de la acuicultura

La acuicultura es una actividad productiva que consiste en la crianza y cultivo de organismos acuáticos como por ejemplo peces, moluscos, crustáceos o plantas (FAO, 2016). Los productos derivados de la acuicultura, son actualmente una de las fuentes más importantes de proteína animal en el mundo, representando un 30% del total de proteína consumida por parte de países en vías de desarrollo y un 15% de la proteína total consumida en Europa y Estados Unidos. La producción acuícola mundial en el año 2015 alcanzó los 199,7 Mt, de las cuales el 49% correspondió a peces, 27.7% algas, 15.5% moluscos, 6.9% crustáceos, mientras que la producción de anfibios y reptiles y otros invertebrados han representado el 0.5% y 0.4%, respectivamente (APROMAR, 2017). Al año 2014, el 88.91% de la producción total se centraba en Asia, siendo China el principal productor, país que a su vez representa más del 60% de la producción acuícola mundial. Otros continentes como África, Europa y Oceanía representan el 2.32%, 3.97% y 0.26%, de la producción acuícola mundial, respectivamente. Por otro lado, América representa el 4.54% de la producción total, en la que Chile por sí solo tiene una participación del 36.34% en relación al resto del continente, posicionándose como el principal país productor acuícola de América (FAO, 2016).

En este contexto, Chile cuenta con condiciones geográficas favorables para el desarrollo de la acuicultura. Actualmente, la actividad acuícola de mayor importancia nacional se centra en la salmonicultura la que ha presentado un desarrollo sostenido durante los últimos años. Dentro de las especies salmonídeas que destacan su producción en Chile se encuentran el salmón de Atlántico (*Salmo salar*), salmón del Pacífico (*Oncorhynchus kisutch*) y la trucha arcoíris (*Oncorhynchus mykiss*), de las cuales a Diciembre del año 2016 se cosecharon 502,400 toneladas, 101,800 toneladas y 71,400 toneladas, respectivamente (SUBPESCA, 2017). El estado sanitario de los peces es uno de los factores de producción más importantes para efectos del retorno económico percibido por la industria del salmón. Es por esto que las enfermedades infecciosas y parasitarias son las patologías que generan mayores pérdidas económicas en los sistemas nacionales de cultivo de peces salmónidos. Por ejemplo, para *S. salar*, *O. kisutch* y *O. mykiss*, las causas de mortalidad infecciosas

durante el año 2016 fueron de 7.7%, 5.3% y 28.2% respectivamente, pero, a su vez dentro de este grupo, la enfermedad Síndrome Rickettsial del Salmón (SRS) es responsable del 74.6%, 13.3% y 86.8% de las mortalidades, respectivamente (SERNAPESCA, 2016).

Mejoramiento genético de salmónidos

Debido a la creciente demanda de proteína, ha surgido la necesidad de mejorar los sistemas de producción animal para generar productos de forma más eficiente y sustentable. Dentro de las medidas que se han debido incorporar para este fin, el mejoramiento genético ha tenido gran impacto. En especies acuícolas, el mejoramiento genético es relativamente reciente en comparación con especies terrestres. En términos generales, los programas de mejoramiento genético tienen por objetivo aumentar la rentabilidad de los sistemas productivos por medio de la crianza selectiva de individuos (Gjedrem, 2012). De acuerdo con esto, todas las características heredables y económicamente relevantes se incluyen dentro del objetivo de selección, como son por ejemplo la tasa de crecimiento, color de filete y la resistencia a enfermedades (Gjedrem, 2012; Villanueva *et al.*, 2011). Para el caso de rasgos asociados al crecimiento, la respuesta a la selección ha sido tan alta como 10% por generación, como es el caso del salmón del Atlántico (Gjoen y Bentsen, 1997; Gjedrem, 2012).

Actualmente, la crianza selectiva de individuos está basada en la selección de los peces de valor genético aditivo superior, los cuales serán los reproductores que darán origen a la próxima generación. El método tradicionalmente utilizado para seleccionar individuos está basado en información fenotípica obtenida desde registros individuales y familiares (padres, madres, hermanos completos, medios hermanos u otros parientes) la que permite estimar los efectos fijos y predecir los efectos aleatorios (valores genéticos aditivos) por medio del método BLUP (*Best Linear Unbiased Prediction*) simultáneamente (Henderson, 1984). Particularmente en el caso de la evaluación genética para resistencia a enfermedades, sólo se pueden utilizar registros desde hermanos completos y medios hermanos de los candidatos a la selección, pues necesariamente se debe recurrir al sacrificio de los individuos una vez finalizada la prueba de desafío experimental, la cual permite obtener los fenotipos de sobrevivencia frente a patógenos específicos. En consecuencia,

para este rasgo en particular se realiza selección familiar, la que sólo explota la mitad de la variación genética para esa característica, ya que no toma en cuenta la variación existente dentro de cada familia (Sonesson y Meuwissen, 2009). Debido a esto, el progreso genético en cada generación está limitado por la menor precisión de valores genéticos estimados al utilizar sólo información desde hermanos completos y medios hermanos de los candidatos a selección, en comparación con la precisión que se podría obtener al considerar la inclusión de información proveniente de estos últimos (Falconer y Mackay, 2001).

Resistencia a enfermedades en peces

Uno de los principales factores a considerar en producción animal, es el estado sanitario de los individuos, el cual afecta directamente a la industria generando grandes pérdidas económicas. El mejoramiento genético se ha posicionado como una estrategia viable y sostenible para prevenir y controlar enfermedades, tendiendo a mejorar la productividad y el bienestar de los animales (Bishop y Woolliams, 2014; Yáñez *et al.*, 2014b).

En este contexto, la resistencia a enfermedades se define como la capacidad del huésped para limitar la infección causada por patógenos, la capacidad de iniciar y mantener respuestas dirigidas a evitar, prevenir el establecimiento de un agente infeccioso y/o eliminarlo del organismo (Bishop y Woolliams, 2014; Raberg *et al.*, 2007). Este rasgo se puede medir como la supervivencia y/o mortalidad de los hermanos completos de candidatos a selección durante un brote infeccioso o prueba de desafío experimental (Odegard *et al.*, 2011; Yáñez *et al.*, 2014c; Yáñez y Martínez, 2010). De acuerdo con esto, los desafíos experimentales bajo ambiente controlado son más apropiados que las pruebas de campo debido a que al controlar las variables ambientales se puede reducir la varianza fenotípica asociada al ambiente y por lo tanto obtener mayores heredabilidades (h^2) para los rasgos en evaluación (Gjoen *et al.*, 1997; Odegard *et al.*, 2011). Conceptualmente, la heredabilidad expresa la proporción de la variación fenotípica total que es atribuible a la variación aditiva de un rasgo en particular (Falconer y Mackay, 2001). Además, en los desafíos experimentales frente a patógenos, es necesario conocer la genealogía de los peces y contar con la identificación individual de éstos, para su posterior utilización en la evaluación genética. Algunas formas en que son expuestos los peces a los organismo

patógenos consideran inoculación por inmersión, cohabitación o inyección intraperitoneal con el objetivo de registrar la sobrevivencia o mortalidad de los individuos (Nordmo *et al.*, 1997).

Entre las enfermedades bacterianas de mayor impacto para el cultivo de peces salmonídeos a nivel nacional, el Síndrome Rickettsial del Salmón (SRS) causada por la bacteria intracelular gram-negativa *Piscirickettsia salmonis*, es uno de los principales patógenos que amenaza la salud de los peces en Chile (Yáñez *et al.*, 2013), la que puede llegar a producir pérdidas económicas de ~25% del beneficio total en salmón de exportación producido en el país (Rozas y Henríquez, 2014). En relación a las enfermedades virales, una de las más relevantes es la Necrosis Pancreática Infecciosa (IPN) (Smith *et al.*, 2001), la cual se presenta principalmente en trucha arcoíris pero también pueden afectar a *S. salar* y *O. kisutch*. Esta enfermedad es altamente contagiosa y puede afectar a los peces tanto en etapas de agua dulce y/o salada (Roberts y Pearson, 2005), reportando mortalidades acumuladas desde ~10% a 90% del total de peces cultivados (Callister y Owens, 1995), en el caso de Chile, esta enfermedad afecta mayormente a *O. kisutch* generando el 31.2% de las mortalidades por causa infecciosa (SERNAPESCA, 2016). Dentro de las enfermedades parasitarias, el crustáceo copépodo *Caligus rogercresseyi* ha generado pérdidas económicas del orden del ~6% del valor total producido por la industria del salmón en el mundo (Costello, 2009). Recientemente, se ha descrito la presencia de variación genética para la resistencia a estas enfermedades, por lo que es viable mejorar este rasgo a través de mejoramiento genético y permitir un mejor control de estas patologías en la industria acuícola nacional (Yáñez *et al.*, 2014b).

Marcadores moleculares

Actualmente, el desarrollo de nuevas tecnologías para obtener mayor cantidad y mejor calidad de información desde el genoma, en conjunto con la progresiva disminución de los costos de obtención de esta información, están cambiando constantemente los escenarios sobre los cuales se puede realizar mejoramiento genético. Dichos avances han abierto la posibilidad de identificar gran cantidad de marcadores moleculares asociados a genes que pueden estar influyendo sobre alguna característica en diversas especies, por ejemplo para

resistencia a enfermedades. La identificación de estos marcadores moleculares ha proporcionado una oportunidad para mejorar la respuesta a la selección de individuos para diversos rasgos de interés económico (Dekkers, 2004).

Dentro de la variedad de marcadores de ADN que existen, destacan los SNPs, los cuales se encuentran ampliamente distribuidos a lo largo de todo el genoma de las especies. Estos corresponden a una modificación en un nucleótido en particular, ubicado en un locus específico en la secuencia de ADN. Este polimorfismo puede representar transiciones, transversiones, inserciones y deleciones en la base nitrogenada (Lander, 1996). Además, este tipo de marcador molecular es de elección para ser utilizado en los métodos de genotipado debido a su abundancia, estabilidad y facilidad para ser analizados (Vignal *et al.*, 2002), por lo que son de gran utilidad para realizar evaluaciones genéticas utilizando información genómica, construcción de mapas genéticos y el mapeo de QTL (*Quantitative trait loci*) mediante GWAS (*Genome-Wide Association Studies*), permitiendo la identificación de genes responsables de la variación genética para determinados caracteres en especies de interés comercial (Hayes y Goddard, 2010). Recientemente el uso de tecnología NGS (*Next Generation Sequencing*), está aumentando en forma exponencial la disponibilidad de información molecular para estos fines. Este método de secuenciación masiva, ha facilitado la detección de marcadores genéticos, los cuales tienen múltiples aplicaciones para el desarrollo del cultivo de especies acuáticas (Yáñez *et al.*, 2014c; Yáñez *et al.*, 2015).

Los marcadores moleculares se han ido incorporando recientemente como herramienta para realizar selección de individuos dentro de programas de mejoramiento genético acuícolas. Utilizando paneles de SNPs se pueden predecir los valores genético-genómicos para diferentes caracteres de relevancia comercial y realizar la selección de individuos con mayor precisión (Odegard *et al.*, 2014). Actualmente, se cuenta con paneles de SNP de alta densidad para diferentes especies de interés comercial como por ejemplo bovinos (Matukumalli *et al.*, 2009), cerdos (Ramos *et al.*, 2009), aves (Groenen *et al.*, 2011), salmón del Atlántico (Houston *et al.*, 2014; Yáñez *et al.*, 2016) y trucha arcoíris (Palti *et al.*, 2015).

Selección genómica

Los recientes avances en relación a tecnologías de secuenciación masiva han contribuido a detectar variantes genéticas involucradas en la variación fenotípica de una amplia gama de rasgos en diferentes especies. Dentro de las aplicaciones que tienen los marcadores moleculares, específicamente en especies terrestres y acuícolas de interés comercial, se encuentra la evaluación genética a través de selección genómica (Goddard y Hayes, 2009). Este método fue propuesto por Meuwissen *et al.*, 2001 como una forma de selección asistida por marcadores moleculares que cubren todo el genoma, con el objetivo de que todos los QTL estén en desequilibrio de ligamiento (LD, *Linkage Disequilibrium*) con al menos un marcador, permitiendo la predicción de los valores genéticos de los animales a partir de éstos. Los SNPs han permitido la implementación de este método para realizar la evaluación genética de los candidatos a selección en función en un carácter de interés comercial en particular (Goddard y Hayes, 2007).

Para aplicar el método denominado GBLUP (*Genomic Best Linear Unbiased Prediction*), en primer lugar es necesario contar con información de genotipos densos de SNPs de un número determinado de individuos. En segundo lugar, se requiere una ecuación de predicción, derivada de los datos moleculares y registros fenotípicos de los individuos de la muestra de entrenamiento, la cual se basa en modelos lineales aplicados para predecir los valores genéticos de los individuos en la muestra de validación o candidatos a la selección (Taylor, 2014). La ventaja de la selección genómica es que no requiere de los registros fenotípicos de los candidatos a selección para evaluarlos. En este sentido, los marcadores moleculares representan una fuente de información adicional, lo que implica una mayor precisión con la que se estiman los valores genéticos y por lo tanto mejorando la respuesta a la selección. Además, este método permite reducir los costos de las pruebas de progenie permitiendo preseleccionar individuos de ambos sexos a temprana edad, posibilitando el genotipado de los individuos desde su nacimiento, lo que reduce los intervalos generacionales, especialmente en especies terrestres (Meuwissen *et al.*, 2001) y más recientemente en especies acuícolas (Banger *et al.*, 2017; Correa *et al.*, 2017a; Vallejos *et al.*, 2017).

En el contexto de los esquema de selección acuícolas, la selección genómica puede ser de gran utilidad ya que los objetivos de mejoramiento incluyen rasgos basados en información fenotípica proveniente de hermanos completos y medios hermanos debido a la imposibilidad de ser medidos en los candidatos a selección sin recurrir al sacrificio de éstos, por ejemplo, resistencia a enfermedades (Villanueva *et al.*, 2011). De acuerdo con esto, la selección genómica permite estimar los efectos de los marcadores y los valores genéticos tanto en los candidatos como en otros individuos emparentados con ellos, resultando en una mayor precisión de selección para esos rasgos, pues se explota tanto la variación genética entre diferentes familias como la variación genética dentro de cada familia (Sonesson y Meuwissen, 2009; Villanueva *et al.*, 2011).

Imputación de genotipos

Una de las principales limitaciones de los métodos de predicción genómica, están asociadas a que la información genotípica de alta resolución requerida para implementar los métodos de evaluación genética mediante GBLUP tienen un costo muy elevado (entre USD \$60 a \$80 por animal). Para mitigar este problema, se pueden utilizar paneles de baja densidad de marcadores moleculares en conjunto con un método denominado “imputación”. Este método usa la estructura de desequilibrio de ligamiento de las regiones del genoma para inferir alelos que no han sido genotipados. Bajo el supuesto de que todas las muestras pertenecen a la misma población, el método considera un set de haplotipos de referencia, el que cuenta con gran cantidad de SNPs genotipados. A partir de éste se completan las posiciones de los genotipos faltantes, haciendo coincidir las bases nitrogenadas del set inicial con cada posición faltante y en caso de que en el set de referencia exista más de una base nitrogenada posible para una cierta posición, el método es optimizado por medio de algoritmos probabilísticos para definir qué base nitrogenada se debe asignar a esa posición (Huang *et al.*, 2012). Existen diferentes métodos en los que se puede basar el proceso de imputación por ejemplo *Hidden Markov Models*, *Likelihood-based diplotype proxies* y *coalescencia*, los cuales se diferencian principalmente en la forma de asignar los haplotipos ancestrales y las probabilidades para determinar los haplotipos faltantes (Halperin y Stephan, 2009). Adicionalmente, en poblaciones que cuentan con información de pedigrí, los factores más importantes que determinan la precisión con que se realiza la imputación

de genotipos de los individuos de la progenie, son la densidad de genotipos de referencia desde paneles de SNPs, idealmente genotipos obtenidos desde paneles de alta densidad y además, el grado de parentesco entre los individuos de referencia y los individuos a imputar, en que los ancestros inmediatos (ambos padres) son los individuos que proveen información para realizar la imputación de genotipos de la progenie con mayor precisión que otros individuos emparentados con ésta (Cleveland y Hickey, 2013; Hickey *et al.*, 2012). Varios programas han sido desarrollados para imputar genotipos, como por ejemplo Beagle (Browning y Browning, 2007), AlphaImpute (Hickey *et al.*, 2012) y FImpute (Sargolzaei *et al.*, 2014) entre otros, los cuales se diferencian principalmente en su precisión de imputación y rapidez de cálculo. Además, es necesario considerar que éstos no están exentos de errores en la imputación (Huang *et al.*, 2012), lo cual dependerá de la estrategia de genotipado utilizada, por ejemplo, el uso de paneles de SNP de baja, media o alta densidad como genotipos de referencia desde individuos teniendo en cuenta el grado de parentesco entre éstos y los individuos a imputar, uso de secuencia completa, selección de marcadores por medio de algún criterio específico (selección aleatoria, espaciada, etc.), eventos como recombinación genética o mutaciones y la cantidad de dichos individuos de referencia disponibles (Huang *et al.*, 2012; Cleveland y Hickey, 2013; Tsai *et al.*, 2017). En este sentido, el uso de paneles de SNPs de baja densidad combinados con estrategias de imputación, se ha propuesto como una alternativa para disminuir los costos de genotipado tanto en especies terrestres (Cleveland y Hickey, 2013) como en especies acuícolas (Tsai *et al.*, 2017).

Estudios de simulación

Las simulaciones por medio de programas específicos son una aproximación regularmente utilizada para investigar sobre la óptima incorporación de información genómica a esquemas de evaluación y selección genética. Dada la dificultad o imposibilidad de realizar análisis empíricos a priori en la práctica, las simulaciones brindan la posibilidad de simular diferentes escenarios para caracteres complejos en los que generalmente se desconoce información acerca de ellos por ejemplo, arquitectura genética de los caracteres de interés o la historia demográfica real de la población en estudio. Por esto, dichos escenarios generalmente tienen en consideración parámetros como arquitectura genética de los

caracteres, número de generaciones de la población, desequilibrio de ligamiento y forma en que se distribuyen los SNP a lo largo del genoma (Nordborg y Tavaré, 2002). Para realizar estas simulaciones, se han desarrollado varios programas computacionales, por ejemplo, AlphaMPSim (Hickey *et al.*, 2014), XSim (Cheng *et al.*, 2015) y QMSim (Sargolzaei y Schenkel, 2009), los cuales requieren definir valores precisos en relación a los parámetros utilizados en las simulaciones para generar escenarios lo más cercanos posible a la realidad, por ejemplo, utilizar valores de heredabilidad descritos en literatura para un carácter específico y alta similitud en niveles de desequilibrio de ligamiento entre los marcadores de datos simulados y reales. Este último punto constituye una dificultad en especies con poca información disponible acerca de su historia demográfica (cuellos de botella recientes, cruzamientos o selección intensiva), como es el caso de la mayoría de las especies acuícolas. En este sentido, herramientas como XSim y QMSim presentan algunas limitaciones. Por ejemplo, XSim no simula loci causales ni implementa evaluación genómica en sus simulaciones, mientras que QMSim requiere de un demandante proceso iterativo inicial (*burning*) para generar patrones realistas de desequilibrio de ligamiento y tampoco realiza evaluación genómica en sus simulaciones. Además, ambos requieren de información previa como input sobre la historia demográfica de la población a estudiar, la cual no siempre está disponible (Pérez-Enciso y Legarra, 2016).

Recientemente se ha desarrollado un programa para simular evaluaciones genéticas que permiten incorporar información molecular llamado “ms2gs” (Pérez-Enciso y Legarra, 2016). Esta herramienta permite evaluar distintos escenarios a partir de información (genotípica y de pedigrí) real o simulada para caracteres complejos y realizar la evaluación genética de éstos. Este algoritmo tiene en cuenta la metodología llamada “*gene-dropping*” para simular los genotipos de las generaciones posteriores a la población fundadora de acuerdo con las leyes de segregación Mendelianas, asumiendo que los genotipos son de carácter bi-alélico (MacCluer *et al.*, 1986). Adicionalmente, “ms2gs” permite definir distintas distribuciones de efectos de QTLs y errores en el genotipado de SNP (Pérez-Enciso y Legarra, 2016).

HIPÓTESIS

Los valores de predicción genómica son más precisos que los valores obtenidos por medio del método tradicional BLUP independiente del error de imputación de genotipos que se considere, en la evaluación genética de trucha arcoíris (*Oncorhynchus mykiss*).

OBJETIVOS

Objetivo General

Determinar y comparar valores de precisión de los valores de cría por medio de método tradicional BLUP y método GBLUP en caracteres de resistencia a enfermedades en trucha arcoíris (*O. mykiss*).

Objetivos Específicos

1. Comparar los valores de precisión para ambas metodologías de evaluación genética (BLUP y GBLUP) a través de cinco generaciones y caracteres con tres niveles de heredabilidad (baja, media y alta) en trucha arcoíris (*O. mykiss*).
2. Comparar el efecto de tres niveles de error de imputación (10%, 7%, y 5%), representando paneles de diferentes densidades de SNP (0.5K, 3K y 7K) versus un panel de alta densidad sobre la precisión en métodos de predicción genómica (GBLUP).

MATERIALES Y MÉTODOS

Software

En el presente estudio se realizaron simulaciones para comparar valores de precisión en la predicción de EBVs (*Estimated Breeding Values*) para rasgos heredables, los cuales deben ser fenotipados mediante pruebas realizadas en hermanos de los candidatos a la selección bajo la estructura de un esquema de selección artificial acuícola. El programa “ms2gs” (Pérez-Enciso y Legarra, 2016) fue utilizado para comparar el método BLUP y GBLUP (selección genómica) por medio de la simulación de diferentes escenarios utilizando una estructura de pedigrí definida y datos genotípicos reales de trucha arcoíris (exclusiva de población fundadora) como “*input*”. El programa permite estudiar caracteres complejos, utilizando “*gene-dropping*” para simular los genotipos y posteriormente calcular los fenotipos de los descendientes a partir de la información inicial para su posterior evaluación por ambos métodos.

Población

Se consideró un esquema de cría de trucha arcoíris con un número de 831 padres y 831 madres como fundadores (F0) de la población completa. Luego, en cada generación se muestrearon aleatoriamente 40 padres y 120 madres (considerando una proporción entre machos y hembras de 1:3, la cual es generalmente utilizada en esquemas de selección en salmónidos) para formar las generaciones siguientes. Cada familia de hermanos completos fue constituida en promedio por 32 individuos al simular F1, F2, F3, F4 y F5, esto generó 3,600 individuos por cada generación y un total de 19,662 individuos al completar la quinta generación (F5).

Genotipado

El ADN fue obtenido desde muestras de aleta pectoral de cada pez utilizando un kit comercial (*DNeasy Blood and Tissue*), luego el ADN de cada muestra fue hibridado en el microarreglo de 57K SNP Affymetrix para trucha arcoíris (Palti *et al.*, 2015). De esta forma, se utilizó la información genotípica disponible desde 1,700 peces, la cual fue utilizada para la posterior simulación como se describe a continuación.

Simulación

A partir de datos genotípicos reales para F0 y considerando la estructura de pedigrí definida previamente para esta población, se simuló los genotipos y fenotipos para las cinco generaciones siguientes. Además, se utilizaron los SNPs que pasaron el control de calidad aplicado a los datos genotípicos desde dicha población fundadora por medio de la herramienta PLINK (Purcell *et al.*, 2007). El control de calidad se llevó a cabo removiendo información faltante desde los genotipos utilizando una tasa de llamada (*call rate*) de 90% para individuos y SNP con el objetivo de evaluar la calidad del ADN de las muestras y la informatividad de los SNP a incluir, respectivamente. Además, individuos con alto nivel de IBD (*Identity By Descent*) fueron removidos utilizando un umbral de 0.05 como valor mínimo, pues representan información duplicada. De acuerdo con este control de calidad, se obtuvo una cantidad total de 42,822 SNP reales disponibles. Las posiciones de los QTLs fueron definidas aleatoriamente en cada cromosoma (4 QTLs por cromosoma, 116 QTLs en total) y sus efectos fueron muestreados desde una distribución gamma con parámetros de forma 0.42 y escala 5.4 (para determinar varianza aditiva) reportada para ganado lechero (Hayes y Goddard, 2001), la que refleja que existe un pequeño número de loci con gran efecto y gran número de loci con pequeño efecto y una distribución gamma con parámetros de forma 0.2 y escala 1 para varianza dominante, lo que refleja un efecto de dominancia pequeño. Se espera una situación similar para caracteres de interés económico en especies acuícolas (Correa *et al.*, 2015; Correa *et al.*, 2017b).

Adicionalmente, paneles de baja densidad fueron simulados usando la herramienta PLINK por medio de su función “*prune*”, la que permite obtener sets de marcadores que se encuentran en desequilibrio de ligamiento unos con otros, los que fueron excluidos del panel de alta densidad. Los archivos correspondientes a estos paneles fueron generados de acuerdo con tres parámetros como se define a continuación:

1. Tamaño de ventana de SNP: 9,425, 150 y 50 SNP para 0.5K, 3K y 7K, respectivamente.
2. Número de SNPs a retirar en cada fase de la ventana: 9,424, 5 y 5 SNP para 0.5K, 3K y 7K, respectivamente.

3. Umbral de VIF (Variance Inflation Factor): 1.5, 1.5 y 2 para 0.5K, 3K y 7K, respectivamente.

De acuerdo con estos parámetros para los paneles de 0.5K, 3K y 7K se consideraron 500, 3,132 y 7,127 SNP como datos iniciales, respectivamente. En este sentido, para comparar el efecto del error de imputación en la evaluación genómica a través de las cinco generaciones y distintos niveles de heredabilidad, se imitó la tasa de error de paneles de genotipado de 0.5K, 3K y 7K SNP simulados utilizando el software “ms2gs”. Para esto, se simuló la imputación de genotipos desde F1 a F5, usando información de SNPs que pueden presentar error de imputación desde archivos creados específicamente para cada panel simulado. Definida en el archivo de parámetros la probabilidad de error de imputación asociada a cada panel de SNP, se asignaron genotipos aleatoriamente erróneos a lo largo de cada panel de izquierda a derecha, respectivamente, para simular dicha tasa de error de imputación.

Escenarios

Los escenarios de simulación variaron de acuerdo con:

1. Error de Imputación: se asumió diferentes tasas de error de imputación en concordancia con las diferentes densidades de los paneles de SNP. Estos porcentaje de error fueron definidos como 10%, 5%, 1% y 0%, los cuales representan la condición más probables asociada a paneles de 0.5K, 3K, 7K y 42K (todos los SNP disponibles), respectivamente (Cleveland y Hickey, 2013). Se definieron los siguientes escenarios para comparar: BLUP, GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K, correspondiendo cada escenario a la utilización de pedigrí y de paneles de 0.5K, 3K, 7K y 42K SNP, respectivamente.
2. Número de generaciones: se consideraron y compararon cinco generaciones consecutivas (1, 2, 3, 4 y 5) de individuos dentro de una población de cría simulada con pedigrí definido a partir de una población fundadora (F0).
3. Heredabilidad (h^2): se utilizaron diferentes niveles de heredabilidad los cuales representan valores reportados en la literatura para caracteres de resistencia a enfermedades en especies salmonídeas. Los valores de heredabilidad utilizados en las simulaciones serán 0.1 (baja), 0.2 (media) y 0.4 (alta).

Se realizaron cinco validaciones cruzadas en cada escenario considerando un 10% aleatorio de la población (excluyendo a los fundadores) como “grupo de entrenamiento” y el 90% restante como “grupo de prueba” en cada validación. Para cada escenario se realizaron 30 réplicas de todo el procedimiento de cálculo. Todas las simulaciones fueron realizadas en un servidor disponible en la Unidad de Genética y Genómica Acuícola de la Facultad de Ciencias Veterinarias y Pecuarias de la Universidad de Chile.

Análisis estadístico

Estimación de valores genéticos y fenotípicos

Se calcularon los valores de cría verdaderos (*True Breeding Value*, TBV) para cada individuo mediante la suma de los valores de efecto de los 116 QTLs en cada carácter. Los valores de cría estimados (*Estimated Breeding Value*, EBV) se obtuvieron por medio de la suma de los efectos de todos los marcadores, como se especifica en la siguiente fórmula:

$$\hat{g}_i = \sum_j^{nSNP} (x_{ij} + \bar{u}_j) \hat{a}_j$$

Donde $nSNP$ es el número de SNPs; x_{ij} es el genotipo para el j -ésimo marcador el que toma valores de -1, 0 y 1 para los genotipos homocigotos para el alelo ancestral, heterocigotos y homocigotos para el alelo derivado, respectivamente para cada SNP; \bar{u}_j es el valor promedio del genotipo y \hat{a}_j es la estimación del efecto sustitución aditivo.

Se simularon los fenotipos mediante la adición de un residuo o varianza ambiental ajustado en función de la heredabilidad que se haya utilizado, definido como $\sigma_e^2 = 1 - (\sigma_a^2 + \sigma_d^2)$, donde σ_a^2 y σ_d^2 corresponden a las varianzas aditivas y dominante, respectivamente. Este cálculo fue realizado de acuerdo a la siguiente fórmula:

$$y_i = \mu + \sum_j^{nSNP} (x_{ij} + \bar{u}_j) a_j + e_i$$

Donde y_i corresponde al fenotipo para el i -ésimo individuo; μ es la media general; $nSNP$ es el número de SNPs; x_{ij} es el genotipo para el j -ésimo marcador el que toma valores de -1, 0 y 1 para los genotipos homocigotos para el alelo ancestral, heterocigotos y homocigotos para el alelo derivado, respectivamente para cada SNP; \bar{u}_j es el valor

promedio del genotipo y a_j es el efecto sustitución aditivo y e_i corresponde al valor residual con distribución normal $\sim N(0, \sigma_e^2)$.

Para la predicción del mérito genético de los peces, se consideró un modelo mixto el cual considera efectos fijos (sexo, familia y edad) y efectos aleatorios (valor genético aditivo para un carácter en particular y el efecto ambiental), como se especifica en la siguiente fórmula:

$$y_i = Xb + Zu + e_i$$

Donde y_i corresponde al fenotipo del carácter a evaluar; X es una matriz de diseño para efectos fijos; b es el vector de efectos fijos; Z es una matriz de diseño para efectos aleatorios; u es un vector de efectos aleatorios, con distribución normal $N\sim(0, G)$ y $var(u) = G = A\sigma_a^2$; e_i es un vector aleatorio (residual) con distribución normal $N\sim(0, R)$ y $var(e) = R = I\sigma_e^2$; G corresponde a la matriz de varianza-covarianza genética aditiva; A es la matriz de parentesco genético; R corresponde a la matriz de varianza-covarianza de los efectos residuales e I corresponde a la matriz identidad.

Se utilizó el modelo tradicional BLUP (Henderson, 1948) para realizar la evaluación genética de los peces teniendo en cuenta información procedente desde sus fenotipos e información de pedigrí simulados, para esto se consideró la siguiente fórmula:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Donde X y Z son matrices de diseño para efectos fijos y efectos aleatorios, respectivamente; X' y Z' son matrices de diseño transpuestas para efectos fijos y efectos aleatorios, respectivamente; A es la matriz de parentesco genético; λ representa un ponderador dependiente de la heredabilidad (h^2); b es un vector de efectos fijos; u es un vector de efectos aleatorios e y corresponde al fenotipo del carácter a evaluar.

Se utilizó el modelo GBLUP (Meuwissen, 2001) para realizar la evaluación genética de los peces a partir de su información molecular desde los genotipos simulados, para esto se consideró la siguiente fórmula:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\lambda \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Donde X y Z son matrices de diseño para efectos fijos y efectos aleatorios, respectivamente; X' y Z' son matrices de diseño transpuestas para efectos fijos y efectos aleatorios, respectivamente; G es la matriz de parentesco genómico; λ representa un ponderador dependiente de la heredabilidad (h^2); b es el vector de efectos fijos; u es un vector de efectos aleatorios e y corresponde al fenotipo del carácter a evaluar.

Estimación de Precisión

Los valores de precisión de los EBVs fueron calculados como:

$$r_{TBV,EBV} = \frac{\sigma_{TBV,EBV}}{\sqrt{\sigma_{TBV}^2 \times \sigma_{EBV}^2}}$$

Donde $r_{TBV,EBV}$ corresponde al coeficiente de correlación de Pearson entre el valor de cría verdadero (TBV) y el valor de cría estimado (EBV), $\sigma_{TBV,EBV}$ corresponde a la covarianza entre los valores de cría verdaderos y estimados, σ_{TBV}^2 corresponde a la varianza de los valores de cría verdaderos y σ_{EBV}^2 corresponde a la varianza de los valores de cría estimados.

La determinación de diferencias significativas para los valores de precisión entre los métodos BLUP y GBLUP para cada escenario se realizó mediante un análisis de varianza (ANOVA) y posteriormente un test de comparaciones múltiples (LSD Fisher) utilizando el software estadístico InfoStat (Di Rienzo *et al.*, 2011).

Estimación de Exactitud

Por medio del mismo software estadístico se realizó un análisis de sesgo de los EBVs obtenidos para la primera (F1) y quinta generación (F5) por medio del coeficiente de regresión lineal (β) calculado entre los TBVs y EBVs como:

$$\beta = \frac{\sum(x - \mu)(y - \tau)}{\sum(x - \mu)^2}$$

Donde β corresponde al coeficiente de regresión lineal entre TBVs y EBVs; x e y corresponden a los TBVs y EBVs, respectivamente; μ y τ representan a los promedios aritméticos de TBVs y EBVs, respectivamente.

Para cada caso se realizó un análisis de varianza (ANOVA) y posteriormente un test de comparaciones múltiples (LSD Fisher), para detectar diferencias significativas entre estos valores.

Evaluación Económica

Con el fin de evaluar la factibilidad económica de incorporar información molecular para seleccionar de manera más precisa peces resistentes a enfermedades de interés comercial, se consideró una empresa productora de ovas de trucha arcoíris. Estas son seleccionadas para resistencia a enfermedades, el volumen total de ovas que se venden cada año es de ~50,000,000 a un precio base de \$200 USD/1,000 ovas seleccionadas utilizando el método BLUP el cual tiene una precisión determinada de los EBVs para caracteres de resistencia dependiendo del valor de heredabilidad de estos. Por otro lado, se evaluó la inclusión de información molecular para realizar selección de las ovas por medio de paneles de SNPs de cuatro densidades distintas (0.5K, 3K, 7K y 42K) en conjunto con la imputación de genotipos en los casos de los tres primeros paneles. El sobreprecio de las ovas obtenidas con cada panel estuvo dado por la diferencia porcentual entre su valor de precisión para estimar los EBVs usando GBLUP y el correspondiente al método BLUP en la quinta generación, considerando la presencia de diferencias significativas ($p < 0.05$) entre estos valores.

A partir de datos técnicos como referencia, los ingresos totales anuales estuvieron dados por la venta de ovas. Además, los costos anuales de esta empresa consideraron la compra de tres licencias para el software “ms2gs” por un monto de \$5,100 USD y un total de salario de \$60,000 USD para tres personas encargadas de realizar la evaluación genética de las ovas. Los costos asociados a paneles de SNPs consideraron los siguientes ítems: 1) Insumos para pruebas de desafío experimental (\$32,000 USD); 2) Reactivos para análisis molecular (extracción de ADN) (\$68,000 USD); 3) Mantenimiento de reproductores (\$8,000,000 USD); 4) Servicio de genotipado Sequenom MassARRAY Iplex Gold e imputación de genotipos (\$100,000 USD para 0.5K SNPs; \$200,000 para 3K SNPs; \$300,000 para 7K SNPs y \$600,000 para 42K SNPs, respectivamente); y 5) Envío de muestras (\$1,000 USD). Los precios de genotipado por muestra se basaron en valores reportados por Huang *et al.*, 2012 y Tsai *et al.*, 2017, por lo que para paneles de 0.5K, 3K, 7K y 42K marcadores incluida la imputación de los genotipos faltantes serán de \$10 USD, \$20 USD, \$30 USD y \$60 USD, respectivamente considerando una población de entrenamiento de 10,000 reproductores. Se calculó la diferencia entre las utilidades percibidas por medio de ambos métodos desde cada panel para determinar el “costo de oportunidad” (utilidad adicional) asociado a la implementación de evaluación genómica versus no implementarla con el fin de determinar el caso más rentable. Adicionalmente, se realizó la evaluación económica para las cuatro primeras generaciones utilizando la misma estructura de costos, ingresos y criterios estadísticos para calcular los sobrepagos de las ovas seleccionadas por medio de GBLUP teniendo en cuenta las diferencias estadísticas entre los valores de precisión reportados para cada método, con la salvedad de que estos resultados no serán reportados gráficamente en este estudio.

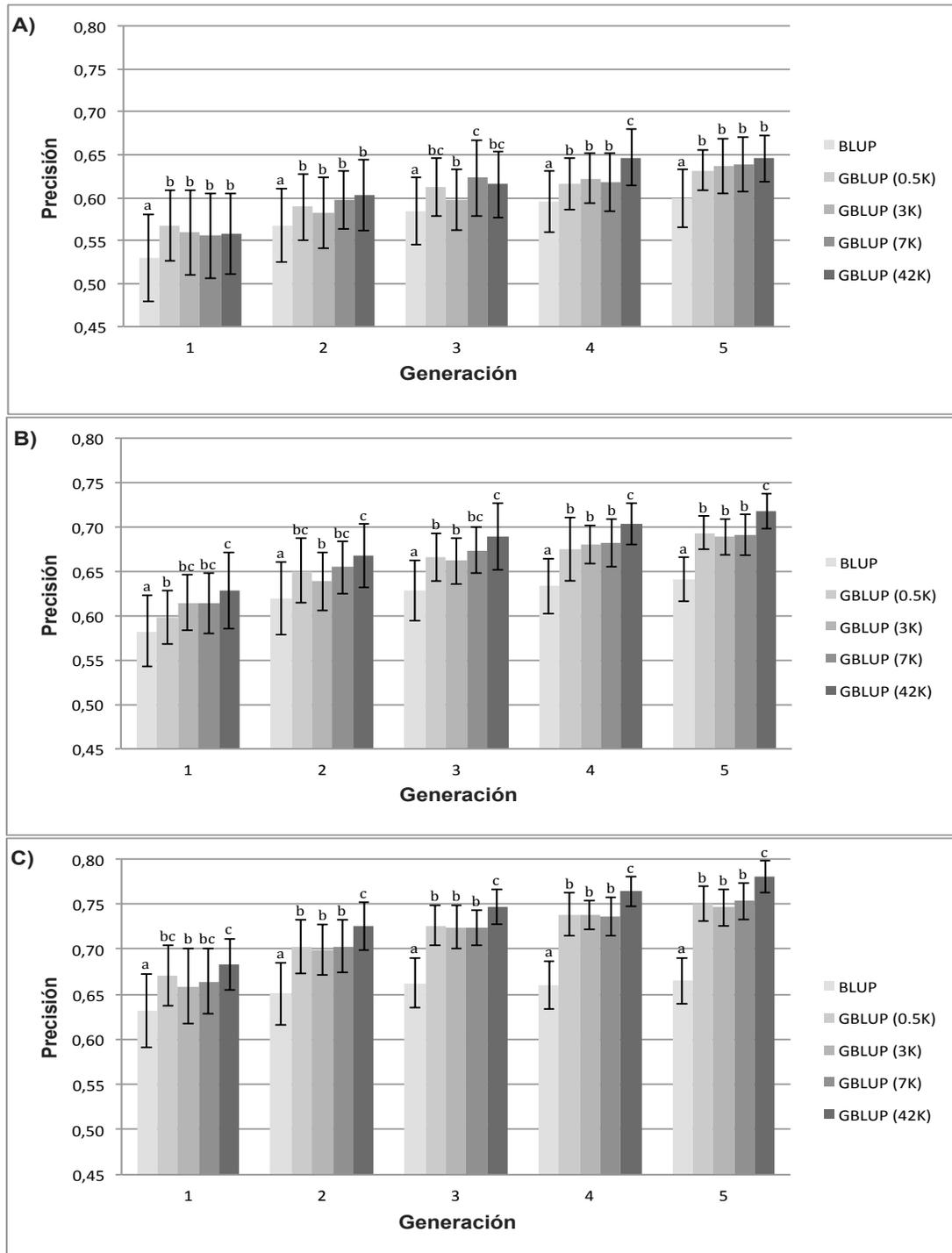
RESULTADOS

BLUP versus GBLUP42K

Los valores de precisión de los EBV por medio de los métodos BLUP y GBLUP utilizando distintas densidades de marcadores en cada escenario son presentados en la **Figura 1**. Como es esperado, los valores de precisión de EBVs tienden a incrementarse cuando se aumenta el número de generaciones y el nivel de heredabilidad. Al comparar BLUP y GBLUP42K a través de cinco generaciones simuladas considerando un carácter de baja heredabilidad ($h^2 = 0.1$), los valores de precisión de EBVs con su respectivo valor de desviación estándar se encuentran entre 0.53 ± 0.05 y 0.60 ± 0.03 y entre 0.56 ± 0.05 y 0.65 ± 0.03 , respectivamente. Para un carácter de media heredabilidad ($h^2 = 0.2$), los valores de precisión de EBVs se encuentran entre 0.58 ± 0.04 y 0.64 ± 0.02 y entre 0.63 ± 0.04 y 0.72 ± 0.02 para BLUP y GBLUP42K, respectivamente y para un carácter de heredabilidad alta ($h^2 = 0.4$) este rango de cambio es entre 0.63 ± 0.04 y 0.67 ± 0.03 y entre 0.68 ± 0.03 and 0.78 ± 0.02 para BLUP y GBLUP42K, respectivamente.

Adicionalmente, los resultados del ANOVA y posterior test de comparaciones múltiples (LSD Fisher) son también presentados en la **Figura 1**. Diferencias significativas ($p < 0.05$) fueron detectadas entre los valores de precisión de EBVs obtenidos por medio del método BLUP y GBLUP42K en todos los escenarios de simulación considerados, a través de las cinco generaciones y los tres niveles de heredabilidad analizados.

Figura 1. Valores de precisión de EBVs en evaluación genómica considerando diferentes densidades de marcadores (GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K) a través de cinco generaciones, tres niveles de heredabilidad A ($h^2 = 0.1$), B ($h^2 = 0.2$) y C ($h^2 = 0.4$), comparadas con valores de precisión obtenidos por medio del método BLUP. Letras distintas (a, b y c) indican diferencias significativas ($p < 0.05$).



Efecto del error de imputación

Para la evaluación genómica utilizando el método GBLUP teniendo en cuenta los paneles de 500 SNP (10% de error), 3K (5% de error) y 7K (1% de error) a través de cinco generaciones y tomando en cuenta un rasgo de baja heredabilidad ($h^2 = 0.1$), los valores de precisión para EBVs se encuentran en un rango de 0.57 ± 0.03 y 0.63 ± 0.02 ; 0.56 ± 0.05 y 0.64 ± 0.03 y entre 0.56 ± 0.05 y 0.64 ± 0.03 , respectivamente. Para un rasgo de mediana heredabilidad ($h^2 = 0.2$) estos valores de precisión se encuentran en un rango entre 0.60 ± 0.03 y 0.69 ± 0.02 ; 0.61 ± 0.03 y 0.69 ± 0.02 y entre 0.61 ± 0.03 y 0.69 ± 0.03 , respectivamente y para un rasgo de alta heredabilidad ($h^2 = 0.4$), el rango de valores de precisión varían entre 0.67 ± 0.03 y 0.75 ± 0.02 ; 0.66 ± 0.04 y 0.75 ± 0.02 y entre 0.66 ± 0.04 y 0.75 ± 0.02 , respectivamente.

Los resultados del ANOVA y posterior test de comparaciones múltiples realizado (LSD Fisher) presentados en la **Figura 1** muestran la presencia de diferencias significativas ($p < 0.05$) sólo entre los valores de precisión obtenidos desde el panel de 7K SNP al ser comparado con los paneles de 500 y 3K SNP, respectivamente en la tercera generación para el rasgo de baja heredabilidad. En el resto de los casos no se detectaron diferencias significativas entre los valores de precisión de EBVs obtenidos desde estos tres paneles intermedios de SNP utilizando el método GBLUP a través de las cinco generaciones y tres niveles de heredabilidad estudiados.

Adicionalmente, en relación a los resultados obtenidos desde el análisis de sesgo para los coeficientes de regresión lineal (β) asociados a TBVs y EBVs en la primera (F1) y quinta generación (F5), el ANOVA y su posterior test de comparaciones múltiples (LSD Fisher) son presentados en la **Tabla 1**, además de su representación gráfica presentada en la **Figura 2.1** y **Figura 2.2**, respectivamente. Para cada nivel de heredabilidad, se consideró incluir a los valores de coeficientes de regresión lineal, el valor de hipótesis nula (H_0) correspondiente a 0 e hipótesis alternativa (H_1) correspondiente a 1, para evaluar la presencia de sesgo en estos resultados.

Tabla 1. Análisis de sesgo para EBVs en la primera (F1) y quinta generación (F5) considerando valores de heredabilidad de 0.1 (baja), 0.2 (media) y 0.4 (alta) por medio de métodos BLUP y GBLUP. Letras distintas (a, b y c) indican diferencias significativas ($p < 0.05$) entre resultados desde test de comparaciones múltiples (LSD Fisher).

Generación	Heredabilidad	Modelo	Precisión	E.E.	Coef. Regresión Lineal	E.E.
F1	0.1	H0	-	-	0.00 A	0.02
		H1	-	-	1.00 B	0.02
		BLUP	0.53	0.01	0.28 C	0.02
		GBLUP (0.5K)	0.57	0.007	0.48 D	0.04
		GBLUP (3K)	0.56	0.009	0.38 E	0.04
		GBLUP (7K)	0.56	0.009	0.43 F	0.04
		GBLUP (42K)	0.56	0.009	0.37 E	0.04
	0.2	H0	-	-	0.00 A	0.02
		H1	-	-	1.00 B	0.02
		BLUP	0.58	0.01	0.37 C	0.02
		GBLUP (0.5K)	0.60	0.005	0.50 D	0.04
		GBLUP (3K)	0.61	0.006	0.50 D	0.04
		GBLUP (7K)	0.61	0.006	0.52 D	0.04
		GBLUP (42K)	0.63	0.08	0.45 D	0.04
	0.4	H0	-	-	0.00 A	0.02
		H1	-	-	1.00 B	0.02
		BLUP	0.63	0.01	0.45 C	0.02
		GBLUP (0.5K)	0.67	0.006	0.56 D	0.04
GBLUP (3K)		0.66	0.008	0.60 E	0.04	
GBLUP (7K)		0.66	0.007	0.52 F	0.04	
GBLUP (42K)		0.68	0.005	0.62 E	0.04	
F5	0.1	H0	-	-	0.00 A	0.01
		H1	-	-	1.00 B	0.01
		BLUP	0.60	0.01	0.37 C	0.01
		GBLUP (0.5K)	0.63	0.004	0.46 D	0.02
		GBLUP (3K)	0.64	0.006	0.44 E	0.02
		GBLUP (7K)	0.64	0.006	0.45 DE	0.02
		GBLUP (42K)	0.65	0.005	0.41 F	0.02
	0.2	H0	-	-	0.00 A	0.01
		H1	-	-	1.00 B	0.01
		BLUP	0.64	0.01	0.46 C	0.01
		GBLUP (0.5K)	0.69	0.003	0.53 D	0.01
		GBLUP (3K)	0.69	0.004	0.56 E	0.01
		GBLUP (7K)	0.69	0.004	0.52 D	0.01
		GBLUP (42K)	0.72	0.004	0.52 D	0.01
	0.4	H0	-	-	0.00 A	0.01
		H1	-	-	1.00 B	0.01
		BLUP	0.67	0.005	0.51 C	0.01
		GBLUP (0.5K)	0.75	0.004	0.60 D	0.02
GBLUP (3K)		0.75	0.004	0.60 D	0.02	
GBLUP (7K)		0.75	0.004	0.57 E	0.02	
GBLUP (42K)		0.78	0.003	0.63 F	0.02	

En todos los casos, los β entre TBVs y EBVs obtenidos por medio de ambos métodos son distintos de $H_0=0$. Se evidenció sesgo en todos los modelos de predicción independiente de la h^2 considerada al comparar F1 y F5, pues todos los valores de sesgo de EBVs usando BLUP y GBLUP independiente de la generación y/o de la h^2 que se considere, muestran ser inferiores a 1, lo que indica una subestimación de los EBVs en este estudio. Además, se reportó que los β asociados a EBVs desde modelos BLUP en cada caso fueron menores que los β asociados a cualquier modelo GBLUP, indistintamente de la h^2 y generación que se tome en cuenta, por lo que el modelo BLUP presenta mayor sesgo de EBVs respecto de GBLUP para selección de individuos. Por otro lado, se evidenció un aumento en los valores de β asociados a EBVs en cada modelo al aumentar el valor de la h^2 tanto en F1 como en F5, lo que indica una dependencia entre el grado de sesgo de EBVs y la h^2 del carácter que se esté estudiando. Esto sugiere que a medida que se aumenta el número de generaciones y/o el valor de la h^2 de un carácter, el nivel de sesgo en los EBVs va disminuyendo y por el contrario, los valores de precisión van aumentando. Al comparar los valores de sesgo de EBVs obtenidos por BLUP y GBLUP42K, es posible notar que tanto para F1 como F5, independientemente del nivel de h^2 considerado, los EBVs obtenidos por BLUP presentan un mayor sesgo que los valores obtenidos desde GBLUP42K. Para el caso de F1, al considerar un valor bajo de h^2 , no se detectaron diferencias significativas en relación al sesgo entre GBLUP3K y GBLUP42K, para un valor medio de h^2 , ocurrió esto al comparar todos los modelos GBLUP y al considerar un valor alto de h^2 , no hubo diferencias significativas entre GBLUP3K y GBLUP42K, respectivamente. Por otra parte, para el caso de F5, al considerar un valor bajo de h^2 , no se detectaron diferencias significativas en relación al sesgo entre GBLUP0.5K y GBLUP7K; y GBLUP3K y GBLUP7K, respectivamente. Para un valor medio de h^2 , no se detectaron diferencias al comparar GBLUP0.5K, GBLUP7K y GBLUP42K, respectivamente, y para un valor alto de h^2 , esto ocurrió al comparar GBLUP0.5K y GBLUP7K. Se reportó en este estudio, que al calcular el cambio porcentual en los valores de precisión entre F1 y F5 para BLUP, GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K para un carácter de baja h^2 se obtuvo, 13%, 11%, 14%, 14% y 16%, respectivamente, para un carácter de h^2 media se reportó, 10%, 15%, 13%, 13% y 14%, respectivamente y para un carácter de alta h^2 se reportó 6%, 12%, 14%, 14% y 15%, respectivamente. Por otra parte, en términos de cambio porcentual en relación

a los coeficientes de regresión lineal entre F1 y F5 para BLUP, GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K para un carácter de baja h^2 se obtuvo 32%, -4%, 16%, 5% y 11%, respectivamente, para un carácter de h^2 media se reportó, 24%, 6%, 12%, 0% y 16%, respectivamente y para un carácter de alta h^2 se reportó, 13%, 7%, 0%, 10% y 2%, respectivamente. Además, al comparar F1 y F5, independientemente del nivel de h^2 que se considere, en términos de cambio porcentual en los valores de precisión con que se estiman los EBVs, se observó que dichos valores usando BLUP, GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K presentan un aumento de 10%, 12%, 14%, 14% y 15%, respectivamente. Por otro lado, al realizar la misma comparación en términos de valores de sesgo de EBVs para estos métodos, se reportaron valores de 23%, 3%, 9%, 5%, 9%, respectivamente. Adicionalmente, si se compara la variación promedio de los valores de precisión y de sesgo de EBVs indistintamente de la generación y nivel de h^2 , se obtienen $\sim 13\%$ y $\sim 10\%$, respectivamente.

Figura 2.1. Representación gráfica de dispersión de TBVs versus EBVs en la primera generación (F1) considerando valores de heredabilidad de 0.1 (baja), 0.2 (media) y 0.4 (alta) por medio de métodos BLUP y GBLUP, con sus respectivas rectas de regresión lineal real (color negro) y recta de regresión ideal ($\beta=1$, color rojo).

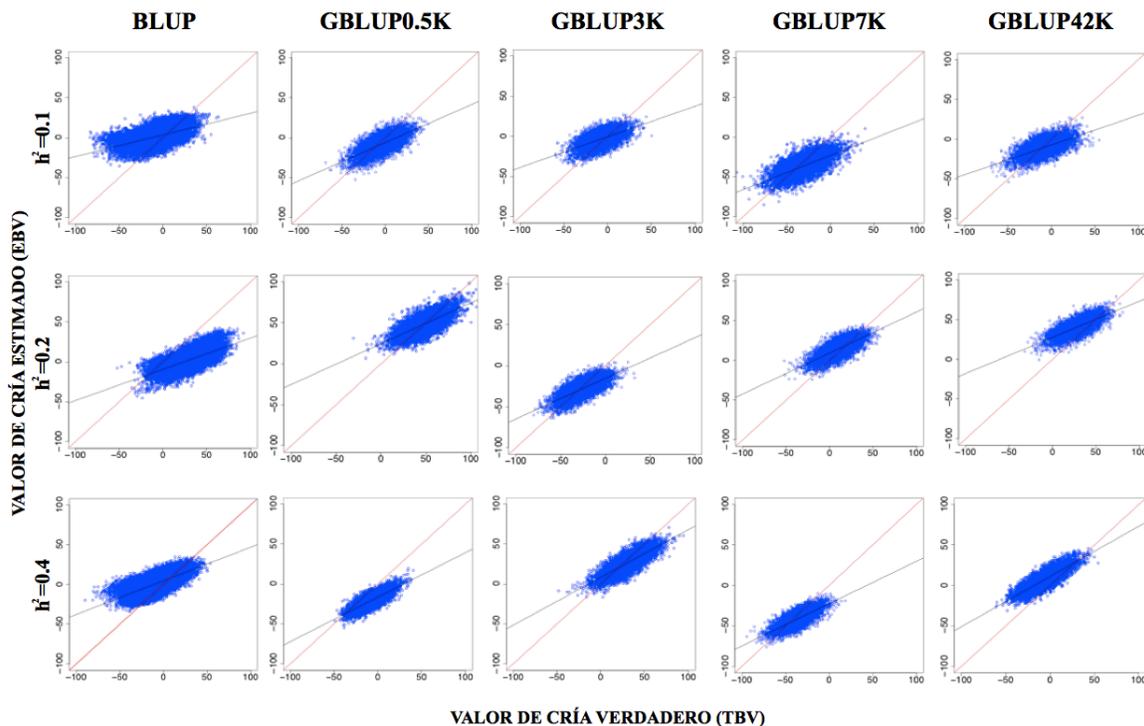
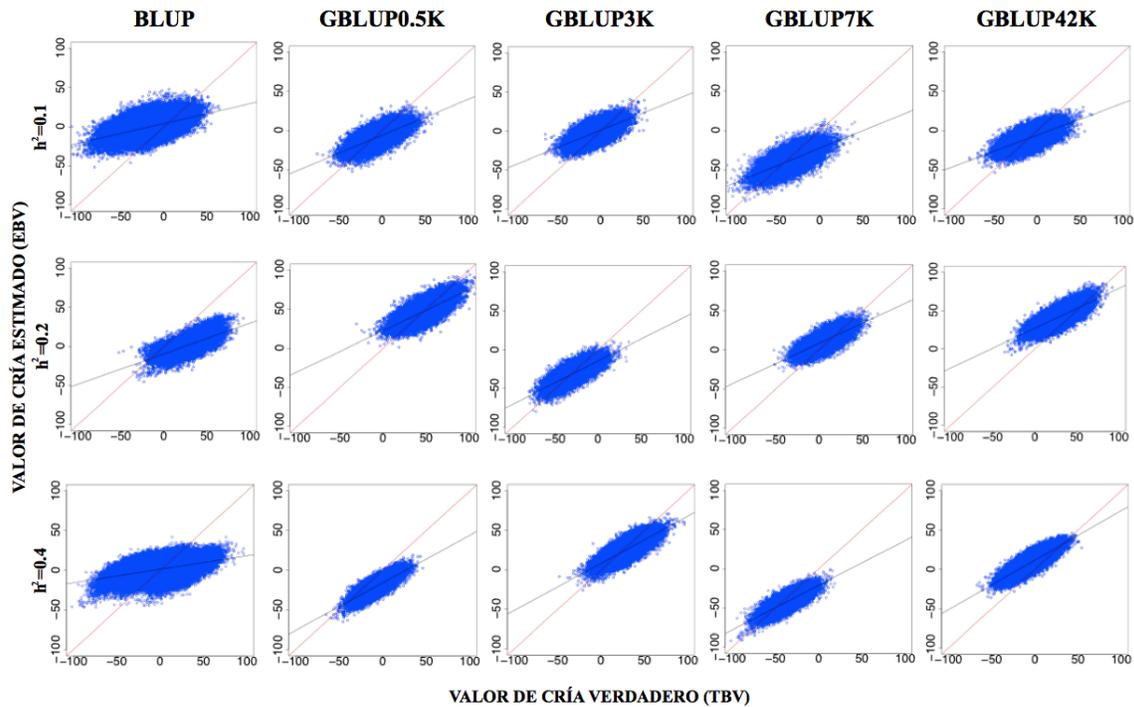


Figura 2.2. Representación gráfica de dispersión de TBVs versus EBVs en la quinta generación (F5) considerando valores de heredabilidad de 0.1 (baja), 0.2 (media) y 0.4 (alta) por medio de métodos BLUP y GBLUP, con sus respectivas rectas de regresión lineal real (color negro) y recta de regresión ideal ($\beta=1$, color rojo).



De acuerdo con los resultados presentados en la **Figura 2.1** y **Figura 2.2**, es posible notar que existe una relación y correlación positiva entre TBVs y EBVs en ambas generaciones. Además, la dispersión de estos valores, es decir, el coeficiente de correlación de Pearson (valor de precisión) tiende a aumentar a medida que aumenta el nivel de heredabilidad. Es posible evidenciar que las rectas de regresión obtenidos se ajustan de mejor manera a la recta teórica ideal con $\beta=1$, en modelos GBLUP en relación al método tradicional BLUP. Estos resultados deben ser complementados con valores reportados en la **Tabla 1**.

Evaluación Económica

Los ingresos, costos, utilidades anuales por producción y costos de oportunidad asociados a mantener BLUP como método de selección de peces versus implementar GBLUP de acuerdo con la venta de ovas seleccionadas para caracteres de resistencia a enfermedades de heredabilidades 0.1, 0.2 y 0.4, las cuales fueron seleccionadas utilizando BLUP, GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K son presentados en la **Tabla 2**.

Tabla 2. Evaluación económica de implementación de selección genómica utilizando cuatro paneles de SNPs de distinta densidad (0.5K, 3K, 7K y 42K) en comparación con evaluación genética basada en información fenotípica durante un año productivo normal.

h^2	Modelo	Precisión F5	Diferencia (%)	Precio 1000 ovas (USD)	Ingreso anual	Costo anual	Utilidad anual	Costo de oportunidad GBLUP-BLUP
0.1	BLUP	0.60	0	200	10,000,000	8,200,000	1,800,000	0
	GBLUP (0.5K)	0.63	6	212	10,600,000	8,266,100	2,333,900	533,900
	GBLUP (3K)	0.64	6	212	10,600,000	8,366,100	2,233,900	433,900
	GBLUP (7K)	0.64	6	212	10,600,000	8,466,100	2,133,900	333,900
	GBLUP (42K)	0.65	6	212	10,600,000	8,766,100	1,833,900	33,900
0.2	BLUP	0.64	0	200	10,000,000	8,200,000	1,800,000	0
	GBLUP (0.5K)	0.69	7	214	10,700,000	8,266,100	2,433,900	633,900
	GBLUP (3K)	0.69	7	214	10,700,000	8,366,100	2,333,900	533,900
	GBLUP (7K)	0.69	7	214	10,700,000	8,466,100	2,233,900	433,900
	GBLUP (42K)	0.72	11	222	11,100,000	8,866,100	2,233,900	433,900
0.4	BLUP	0.67	0	200	10,000,000	8,200,000	1,800,000	0
	GBLUP (0.5K)	0.75	11	222	11,100,000	8,266,100	2,833,900	1,033,900
	GBLUP (3K)	0.75	11	222	11,100,000	8,366,100	2,733,900	933,900
	GBLUP (7K)	0.75	11	222	11,100,000	8,466,100	2,633,900	833,900
	GBLUP (42K)	0.78	14	228	11,400,000	8,866,100	2,533,900	733,900

De acuerdo con estos resultados, para las generaciones F1, F2, F3, F4 y F5 la evaluación genómica utilizando un panel de 0.5K SNPs figuró como la opción más rentable pues fue el modelo que generó las utilidades y costos de oportunidad más altos para la situación planteada, respectivamente. En términos de costo de oportunidad, para valores de heredabilidad baja ($h^2=0.1$) en el caso de la primera generación se reportaron USD 633,900 y para valores de heredabilidad altos ($h^2=0.4$) se reportaron en F2 USD 733,900; F3 USD 1,033,900 y F4 USD 1,133,900; respectivamente. Adicionalmente, de acuerdo a los resultados reportados en la **Tabla 2**, es posible evidenciar que este modelo (GBLUP0.5K) es la opción más rentable indistintamente del nivel de heredabilidad que se considere para dicha evaluación.

DISCUSIÓN

En el presente estudio, se evaluó la precisión con que se obtienen los valores de cría estimados (EBV) por medio del método tradicional basado en información fenotípica y de pedigrí (BLUP) y el método de evaluación genética que tiene en cuenta información de los genotipos de los candidatos a la selección (GBLUP). Se imitó el uso de paneles de baja densidad de SNP considerando para cada uno, diferentes niveles de error de imputación asociado a ellos, a través de cinco generaciones y caracteres de baja, media y alta heredabilidad por medio de la simulación de un esquema de selección acuícola.

Los valores de precisión obtenidos desde GBLUP0.5K, GBLUP3K, GBLU7K y GBLUP42K fueron significativamente mayores que los valores obtenidos por medio del método BLUP a través de todos los escenarios de simulación. Estos resultados indican que la respuesta a la selección en esquemas de selección acuícola puede mejorar utilizando la selección genómica mediante la incorporación de paneles de SNPs densos, como ya ha sido reportado en estudios de simulación previos (Sonesson y Meuwissen, 2009; Lillehammer *et al.*, 2013; Nirea *et al.*, 2012; Vela-Avituá *et al.*, 2015) y en estudios empíricos en otras especies acuícolas (Bangera *et al.*, 2017; Correa *et al.*, 2017a; Tsai *et al.*, 2016; Tsai *et al.*, 2017; Vallejo *et al.*, 2016; Vallejo *et al.*, 2017). Muir, 2007 sugiere que es necesario más de una generación para obtener valores de precisión por medio de GBLUP, mayores que los que se podrían obtener por medio de BLUP. En contraste con esto, en nuestro estudio encontramos diferencias significativas entre los valores de precisión obtenidos por BLUP y GBLUP0.5K, GBLUP3K, GBLU7K y GBLUP42K utilizando sólo una generación teniendo en cuenta los tres niveles de heredabilidad. En ese sentido, Sonesson y Meuwissen, 2009 y Pérez-Enciso *et al.*, 2016 describen que otro factor importante para detectar diferencias entre los valores de precisión de EBVs por medio de ambos métodos, es el tamaño poblacional, es decir, mientras mayor sea la población de entrenamiento, mayores valores de precisión de EBVs se obtendrán. Por lo tanto, el trabajo con especies acuícolas representa una ventaja en este sentido debido a la alta disponibilidad de individuos por familia que se pueden conseguir, lo que se refleja en nuestro resultados al utilizar diferente número de generaciones para comparar ambos métodos. Por otro lado, Nirea *et al.*, 2012 reportó valores de precisión de EBVs de 0.476, 0.571 y 0.720, utilizando

un esquema de selección similar al utilizado en este estudio, a través de dos generaciones formadas por cruzamiento aleatorio con valores de heredabilidad de 0.05, 0.1 y 0.4, respectivamente, pero en este estudio no se consideró la densidad de SNP ni una estrategia de imputación como fuente de variación. Otro estudio realizado por Vela-Avituá *et al.*, 2015 en base a simulaciones, reportó que los valores de precisión de EBV utilizando selección genómica considerando un panel de genotipado de alta densidad (400K SNP) fueron mayores que los valores de precisión obtenidos por medio de BLUP, reportando una diferencia aproximada de un 20% entre ambos métodos, pero sin considerar una estrategia de imputación de genotipos en el análisis. Un estudio similar realizado por Tsai *et al.*, 2017 en salmón del Atlántico, considerando dos densidades de marcadores (0.25K y 25K), imputación de genotipos, dos rasgos (resistencia contra piojo de mar y peso corporal), dos generaciones y validación cruzada, los valores de precisión de EBVs para ambos rasgos fueron marginalmente menores que los valores obtenidos utilizando los genotipos verdaderos (0.58 versus 0.60 y 0.67 versus 0.69, para resistencia a piojo de mar y peso corporal, respectivamente). Además, estos resultados fueron sustancialmente mayores que los valores de precisión obtenidos por medio del método BLUP, 0.48 y 0.58, para ambos rasgos respectivamente.

Adicionalmente, algunos estudios han reportado el efecto de la densidad de paneles de SNP sobre los valores de precisión de EBV obtenidos por medio de GBLUP, los cuales son consistentes con los resultados obtenidos en el presente estudio, Sonesson and Meuwissen, 2009 reportaron que al utilizar densidades de SNP de 1K, 5K y 10K sin considerar una estrategia de imputación de genotipos para estos paneles, la precisión de EBVs es 0.661, 0.823 y 0.842, respectivamente en un esquema de selección acuícola similar al utilizado en este estudio. En otro estudio realizado por Correa *et al.*, 2017a sobre resistencia a piojo de mar en salmón del Atlántico, al utilizar paneles de 0.5K, 1K, 10K y 25K de densidad de SNP pero sin considerar alguna estrategia de imputación de genotipos, los valores de precisión de EBVs obtenidos por medio de GBLUP fueron 0.45, 0.48, 0.50, 0.50 y 0.50, respectivamente. Resultados similares fueron reportados por Tsai *et al.*, 2016 al comparar dos poblaciones de salmón del Atlántico utilizando seis paneles de SNP de densidades crecientes (0.5K, 1K, 5K, 10K, 20K y 33K), obteniendo valores de precisión de EBV que

aumentan en ~10% desde 0.5K a 33K SNP en un esquema de selección aleatorio de individuos (modelo neutral de selección). Además, un estudio realizado por Tsai *et al.*, 2017, en que se considera una estrategia de imputación de genotipos para el rasgo de resistencia a piojo de mar ($h^2 = 0.22$) y peso corporal ($h^2 = 0.50$) en salmón del Atlántico utilizando paneles de genotipado de 0.25K, 25K e imputación de genotipos y 25K SNP, reportó valores de precisión de EBV para resistencia a piojo de mar de ~0.40, 0.58 y 0.60 y para peso corporal de 0.58, 0.68 y 0.70, respectivamente.

De acuerdo con la literatura, la heredabilidad tiene un gran efecto sobre la precisión con que se obtienen los valores de precisión de EBV obtenidos por medio de BLUP y GBLUP, Nielsen *et al.*, 2009 determinaron que los valores de precisión de EBV aumentan cuando la heredabilidad de los rasgos aumenta de 0.2 a 0.4, lo que se corresponde con los resultados obtenidos por Kolbehdari *et al.*, 2007 y consistentemente con lo que se reportó en nuestros resultados. También Sonesson and Meuwissen, 2009 reportaron que al aumentar la heredabilidad de 0.1 a 0.4, los valores de precisión de EBVs en la quinta generación aumenta de 0.742 a 0.823 en un esquema de selección acuícola que considera 30 hermanos completos por familia. En el presente estudio, al considerar el método GBLUP, un panel de 42K SNP, un esquema de 32 individuos por familia, cinco generaciones y valores de heredabilidad para cada rasgo de 0.1, 0.2 y 0.4, obtuvimos valores similares de precisión de EBVs de 0.65 ± 0.05 , 0.72 ± 0.04 y 0.78 ± 0.03 , respectivamente en relación a los estudios anteriormente mencionados. Además, para un rasgo de baja heredabilidad ($h^2 = 0.1$), en el presente estudio, cinco generaciones no fueron suficiente para evidenciar diferencias entre los valores de precisión de EBVs entre GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K. Sin embargo para rasgos de mediana ($h^2 = 0.2$) y alta heredabilidad ($h^2 = 0.4$), cuatro y dos generaciones fueron suficientes, respectivamente para identificar diferencias entre dichos modelos. En términos generales, estudios que comparan precisión de EBVs, describen un incremento no lineal de estos valores a través de paneles de diferente densidad de SNP y niveles de heredabilidad (Pérez-Enciso *et al.*, 2015), lo cual se condice con los resultados obtenidos en este estudio.

Estudios previos en cerdos y salmón del Atlántico en los cuales se estudia la relación existente entre la densidad de los paneles de SNP utilizados, el nivel de parentesco entre los individuos y el error de imputación asociado a ellos, fueron considerados como referencia para nuestro estudio, Cleveland y Hickey, 2013 reportaron que al considerar padres genotipados de *Sus scrofa*, con un panel de mediana densidad (60K), al genotipar la progenie utilizando paneles de densidad 384 SNP, 768 SNP, 1.5K, 3K y 6K considerando imputación de los genotipos, la precisión de imputación de estos fue 0.944, 0.966, 0.981, 0.99 y 0.994, respectivamente. Recientemente un estudio realizado por Yoshida *et al.*, submitted en salmón del Atlántico, en que se genotiparon padres usando un panel de densidad 50K y la progenie con paneles de densidad 0.5K, 3K y 6K considerando imputación de los genotipos, los valores de precisión de imputación reportados fueron 0.7948, 0.9247 y 0.9457, respectivamente. Teniendo en cuenta estos antecedentes y considerando la falta de información previa reportada en la literatura para trucha arcoíris, se consensuó definir los errores de imputación como 10%, 5% y 1% para paneles de 0.5K, 3K y 7K, respectivamente, con el fin de compensar fuentes de sesgo como eventos de recombinación, mutaciones, baja cantidad de individuos de referencia o tamaño y complejidad del genoma de esta especie de la cual no se han reportado resultados en estos temas.

Por otro lado, Pérez-Enciso *et al.*, 2016 tomando en cuenta un modelo neutral de selección (selección aleatoria de individuos) y otro en que se seleccionan individuos de acuerdo a su ranking de valores genéticos aditivos para selección genómica, reportaron que no existen diferencias en los valores de precisión de EBVs entre ambos modelos. Considerando que en este estudio se utilizó un modelo neutral de selección de individuos, se esperarían resultados similares si se considera la selección de individuos. Adicionalmente, en relación al análisis de sesgo realizado en el presente trabajo, el estudio realizado por Wolc *et al.*, 2011 demostró la presencia de sesgo en los EBVs para GBLUP utilizando regresión lineal entre TBVs y EBVs, reportando coeficientes de regresión lineal de 0.9 y 0.8 para BLUP y GBLUP, respectivamente. Nuestros resultados presentados en la **Tabla 1** muestran valores de coeficientes de regresión lineal (β) inferiores (menores que 1) para ambos modelos considerando heredabilidades de 0.1, 0.2 y 0.4, lo que sugiere una subestimación de los

EBVs, pero, a medida que aumenta la heredabilidad este nivel de sesgo va disminuyendo, lo que se evidenció en una dependencia entre las generaciones, el nivel de heredabilidad utilizado y el nivel de sesgo alcanzado. De hecho, al considerar una generación y un nivel bajo de heredabilidad, se obtienen EBVs sesgados para ambos modelos, pero a medida que se aumenta el valor de la heredabilidad de un carácter, el nivel de sesgo de estos valores disminuye, siendo los modelos GBLUP los que se logran menores valores de sesgo, la misma situación ocurre al considerar cinco generaciones. Por lo que nuestros resultados reportan que en general, los modelos de predicción genómicos a pesar de estar influenciados por el nivel de heredabilidad y/o las generaciones del carácter que se esté evaluando, muestran valores de cría estimados más exactos que los obtenidos por el método tradicional BLUP. De acuerdo con Lynch y Walsh, 1998, la influencia de la heredabilidad sobre el sesgo de EBVs, puede estar relacionada con el hecho de que a medida que aumenta la heredabilidad, los EBVs están más influenciados por el fenotipo, debido a que teóricamente cuando la heredabilidad alcanza un valor de 1, el fenotipo es igual al EBV. Adicionalmente, la presencia de sesgo podría estar relacionada al hecho de que tanto el modelo BLUP como GBLUP asumen una distribución normal para los valores de efectos de QTLs, en circunstancias que en este estudio se está asumiendo una distribución gamma para calcular dichos valores, de acuerdo a lo descrito en la literatura para especies terrestres y acuícolas. Sin embargo, la presencia de sesgo podría explicarse además por discrepancias entre los supuestos del modelo de predicción y la historia de selección de la población de entrenamiento, ya que estos modelos predictivos asumen un modelo neutral de selección (Vitezica *et al.*, 2011), tal como se asume en este estudio, por lo que no representaría una fuente de sesgo. Adicionalmente, se ha reportado en la literatura que existe independencia en el uso de diferentes distribuciones de efectos de QTLs para realizar GBLUP (Shirali *et al.*, 2015; Wientjes *et al.*, 2014; Zhang *et al.*, 2015) e incluso se ha demostrado independencia en el uso de diferentes valores para escala y forma de la distribución gamma utilizada (Hayes y Goddard, 2001).

Por otro lado, la información genómica brinda un alto potencial de incrementar la respuesta a la selección. Sin embargo, los costos de adquirir o desarrollar paneles de genotipado son aún altos para la implementación práctica de la selección genómica en acuicultura. En este

contexto, la imputación de genotipos puede contribuir a la disminución de estos costos, representando una alternativa para la actual implementación de información proveniente desde genotipos en la evaluación genética aplicada a la industria acuícola. El estudio realizado por Huang *et al.*, 2012 reportó que los costos de genotipado por individuo desde un panel de SNP de mediana densidad (60K) y paneles de baja densidad (6K, 3K y 384 SNP) fueron de USD \$120, \$48, \$35 y \$20, respectivamente y que la mayoría de la información desde el panel de alta densidad puede ser recuperada utilizando un panel de SNPs de bajo costo combinado con estrategias de imputación de genotipos, Tsai *et al.*, 2017 respalda esta idea, pues en este estudio se reportó que el costo por muestra de genotipado para un panel de SNPs de mediana densidad (25K), es mayor que el costo de genotipado para un panel de baja densidad (250 SNP), asumiendo un valor estimado de £40 y £5 por muestra, respectivamente. Además, el costo total del genotipado para selección genómica considerando una estrategia de imputación de los genotipos es alrededor de un 60% menor que genotipar todas las muestras utilizando un panel de mediana densidad de SNPs.

En este sentido, los resultados reportados a partir de la evaluación económica, para evaluar la factibilidad de implementar el método GBLUP como forma de mejorar la precisión con que se seleccionan individuos en programas de mejoramiento genético acuícolas, sugieren que, los ingresos, costos, utilidades anuales y costos de oportunidad por concepto de producción y venta de ovas seleccionadas para caracteres de resistencia a enfermedades de interés comercial con valores de heredabilidad de 0.1, 0.2 y 0.4, las cuales fueron seleccionadas utilizando BLUP, GBLUP0.5K, GBLUP3K, GBLUP7K y GBLUP42K presentados en la **Tabla 2**. Dichos resultados, sugieren que la implementación de un panel de 500 SNPs en conjunto con la imputación de genotipos es la opción más rentable, indistintamente de la generación que se considere pero se evidenció que existe una relación entre el nivel de heredabilidad y la generación pues para la primera generación el modelo óptimo que maximiza la utilidad considera el método GBLUP0.5 para caracteres de baja heredabilidad y para las cuatro generaciones restantes, el modelo óptimo considera el método GBLUP0.5K para caracteres de alta heredabilidad. Esto teniendo en consideración que el criterio principal de decisión son los valores de precisión de EBVs obtenidos en cada

caso, los que de acuerdo a nuestros resultados reportaron aumentar (~13%) en circunstancias que el nivel de sesgo de EBVs se mantiene relativamente constante (~0%) a través de las generaciones para un nivel de heredabilidad y modelo a utilizar específicos.

Un aspecto relevante es la presencia o ausencia de diferencias significativas en los valores de precisión de EBVs en los escenarios que imitan el uso de paneles de SNP de densidad 0.5, 3K y 7K. Por lo tanto, el presente estudio sugiere que un panel de 500 SNP (considerando un 10% de error de imputación) podría ser una opción a bajo costo para incorporar información molecular en la evaluación genética de individuos dentro de un esquema de selección de *O. mykiss* para caracteres de resistencia de heredabilidad baja (0.1), media (0.2) o alta (0.4). En este sentido, las implicancias de este estudio para los esquemas de selección acuícola están principalmente ligadas con la potencial implementación de dicho panel de baja densidad combinado con estrategias de imputación de genotipos para selección genómica.

Por lo tanto, nuestros resultados respaldan el hecho de que en esquemas de selección utilizados en acuicultura y específicamente para el caso de la trucha arcoíris (*Oncorhynchus mykiss*), la selección genómica puede aumentar la precisión con que se seleccionan individuos en comparación con el método BLUP. Consecuentemente, empresas relacionadas a la acuicultura que poseen programas de mejoramiento genético se podrían ver beneficiadas con la implementación de este método aplicado a caracteres que son imposibles de medir en los candidatos a ser seleccionados, como es el caso de los caracteres de resistencia a enfermedades. Esto representa una ganancia en términos económicos por concepto de utilidades reportadas producto de una selección más precisa de individuos resistentes a enfermedades de relevancia comercial, pero también en términos de tiempo, pues se podrían evaluar peces desde que son ovas, evitando tener que esperar a que sean aptos para ser desafiados experimentalmente, con todos los costos asociados que esto implica.

CONCLUSIONES

De acuerdo con los resultados obtenidos en el presente estudio desde simulaciones a partir de genotipos reales de *O. mykiss*, fue posible determinar que:

1. Los valores de predicción usando GBLUP son más precisos que los valores obtenidos por BLUP independiente de la heredabilidad y número de generaciones utilizado.
2. Se evidenció un aumento en la precisión de EBVs a través de las cinco generaciones y tres niveles de heredabilidad estudiados para ambos métodos de predicción.
3. Los valores de precisión de EBVs desde paneles de SNP de baja densidad (0.5K, 3K y 7K SNPs) y error de imputación (10%, 5% y 1%, respectivamente) considerado, mostraron ser similares estadísticamente en todos los casos.
4. GBLUP reportó ser un método de predicción más exacto que el método tradicional BLUP al comparar EBVs en F1 y F5, independientemente del valor de heredabilidad considerado.
5. En general, se reportó una disminución en el nivel de sesgo, es decir, un aumento en la exactitud en los EBVs obtenidos desde ambos métodos de predicción al aumentar el número de generaciones y/o el valor de heredabilidad considerado.
6. En términos generales, la precisión de EBVs fue el valor que reportó los mayores aumentos porcentuales a través de las generaciones y niveles de heredabilidad, en comparación con los aumentos porcentuales de los coeficientes de regresión lineal de EBVs.
7. La simulación de un panel de baja densidad (500 SNPs) considerando imputación de los genotipos reportó ser la opción más rentable para realizar selección genómica independiente de la heredabilidad del carácter evaluado.
8. Finalmente, a modo de recomendación, este trabajo reportó que la utilización de un modelo GBLUP0.5K para selección de individuos maximiza las utilidades monetarias, cuando se considera una sola generación y caracteres de baja heredabilidad, y si se consideran dos, tres, cuatro o cinco generaciones, se deberían evaluar caracteres de heredabilidad alta.

BIBLIOGRAFÍA

- ASOCIACIÓN EMPRESARIAL DE PRODUCTORES DE CULTIVOS MARINOS DE ESPAÑA (APROMAR).** 2017. Informe de la acuicultura Española, APROMAR-ESACUA-OPP. España. 93 p.
- BANGERA, R.; CORREA, K.; LHORENTE, J.; FIGUEROA, R.; YÁÑEZ, J.** 2017. Genomic predictions for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). BMC Genomics. 18: 121-132.
- BISHOP, S.; WOOLLIAMS, J.** 2014. Genomics and disease resistance studies in livestock. Liv. Sci. 166: 190-198.
- BROWNING, S.; BROWNING B.** 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81(5): 1084-1097.
- CALLISTER, P.; OWENS, W.** 1995. Assessment of the virulence of fish and molluscan isolates of infectious pancreatic necrosis virus for salmonid fish by challenge of brook trout (*Salvelinus fontinalis*). J. Fish. Dis. 18(1): 97-103.
- CHENG, K.; GARRICK, D.; FERNANDO, R.** 2015. XSim: simulation of descendants from ancestor with sequence data. G3. 5(7): 1415-1417.
- CLEVELAND, M.; HICKEY, J.** 2013. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. J. Anim. Sci. 91: 3583-3592.
- CORREA, K.; LHORENTE, J.; LÓPEZ, M.; BASSINI, L.; NASWA, S.; DEEB, N.; DI GENOVA, A.; MAAS, A.; DAVIDSON, W.; YÁÑEZ, J.** 2015. Genome-wide association with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar L.*) chromosomes. BMC Genomics. DOI 10.1186/s12864-015-2038-7.
- CORREA, K.; BANGERA, R.; FIGUEROA, R.; LHORENTE, J.; YÁÑEZ, J.** 2017a. The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*), Genet. Sel. Evol. 49: 15-19.

- CORREA, K.; LHORENTE, J.; BASSINI, L.; LÓPEZ, M.; DI GENOVA, A.; MAAS, A.; DAVIDSON, W.; YÁÑEZ, J.** 2017b. Genome wide association study for resistance to *Caligus rogercresseyi* in Atlantic salmon (*Salmo salar L.*) using a 50K SNP genotyping array. *Aquaculture*. 472: 61-65.
- COSTELLO, M.** 2009. The global economic cost of sea lice to the salmonid farming industry. *J. Fish Dis.* 32: 115-118.
- DEKKERS, J.** 2004. Commercial application of marker- and gene- assisted selection in livestock: Strategies and lessons. *J. Anim. Sc.* 82(13): 313-328.
- DI RIENZO, J.; CASANOVES, F.; BALZARINI, M.; GONZALEZ, L.; TABLADA, M.; ROBLEDO, Y.** 2011. InfoStat versión 2011. Grupo InfoStat. FCA, Universidad Nacional de Córdoba. Argentina. 8: 195-199.
- FALCONER, D.; MACKAY, T.** 2001. Introduction to quantitative genetics. 4th ed. Essex. UK. 386 p.
- FAO.** 2016. Estado mundial de la pesca y la acuicultura: Contribución a la seguridad alimentaria y la nutrición para todos. Roma. Italia. 224 p.
- GJEDREM, T.** 2012. Genetic improvement for the development of efficient global aquaculture: A personal opinion review. *Aquaculture*. 349: 12-22.
- GJOEN, H.; BENTSEN, H.** 1997. Past, present and future of genetic improvement in salmon aquaculture. *J. Marine. Sc.* 54(6): 1009-1014.
- GJOEN, H.; REFSTIE, T.; ULLA, O.; GJERDE, B.** 1997. Genetic correlations between survival of Atlantic salmon in challenge and field tests. *Aquaculture*. 158(3-4): 277-288.
- GODDARD, M.; HAYES, B.** 2007. Genomic selection. *J. An. Breed. Gen.* 124(6): 323-330.
- GODDARD, M.; HAYES, B.** 2009. Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nat. Rev. Gen.* 10: 381-391.
- GROENEN, M.; MEGENS, H.; ZARE, Y.; WESLEY, W.; LADEANA, H.; CROOIJMANS, R.; VEREIJKEN, A.; OKIMOTO, R.; MUIR, W.; CHENG, H.** 2011. The development and characterization of a 60K SNP chip for chicken. *BMC Genomics*. 12: 274-282.

- HALPERIN, E.; STEPHAN, D.** 2009. SNP imputation in association studies. *Nat. Biotech.* 27(4): 349-351.
- HAYES, B.; GODDARD, M.** 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* 33(2001): 209-229.
- HAYES, B.; GODDARD, M.** 2010. Genome-wide association and genomic selection in animal breeding. *Genome.* 53: 876-883.
- HAYES, B.; BOWMAN, P.; CHAMBERLAIN, A.; GODDARD, M.** 2009. Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* 92(2): 433-443.
- HENDERSON, C.** 1984. *Applications of Linear Models in Animal Breeding.* Guelph University. Press. Guelph. Canada. 462 p.
- HICKEY, J.; KINGHORN, B.; TIER, B.; VAN DER WERF, J.; CLEVELAND, M.** 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation method. *Genet. Sel. Evol.* 44: 9-20.
- HICKEY, J.; GORJANC, G.; HEARNE, S.; HUANG, B.** 2014. AlphaMPSim: flexible simulation of multi-parent crosses. *Bioinformatics.* 30(18): 2686-2688.
- HOUSTON, R.; TAGGART, J.; CÉZARD, T.; BEKAERT, M.; LOWE, N.; DOWNING, A.; TALBOT, R.; BISHOP, S.; ARCHIBALD, A.; BRON, J.; PENMAN, D.; DAVASSI, A.; BREW, F.; TINCH, A.; GHARBI, K.; HAMILTON, A.** 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics.* 15: 90-102.
- HUANG, Y.; HICKEY, J.; CLEVELAND, M.; MALTECCA, C.** 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet. Sel. Evol.* 44: 25-32.
- KOLBEHDARI, D.; SCHAEFFER, L.; ROBINSON, J.** 2007. Estimation of genome-wide haplotype effects in half-sibs designs. *J. Anim. Breed. Genet.* 124(6): 356-361.
- LANDER, E.** 1996. The new genomics: global views of biology. *Science.* 274: 536-539.
- LILLEHAMMER, M.; MEUWISSEN, T.; SONESSON, A.** 2013. A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45: 39-46.
- LYNCH, M; WALSH, B.** 1998. *Genetics and analysis of quantitative traits.* Sinauer Press. Sunderland, USA. 980 p.

- MACCLUER, J.; VANDEBERG, J.; READ, B.; RYDER, O.** 1986. Pedigree analysis by computer simulation. *Zoo Bio.* 5: 147-160.
- MATUKUMALLI, L.; LAWLEY, C.; SCHNABEL, R.; TAYLOR, J.; ALLAN, M.; HEATON, M.; O'CONNELL, J.; MOORE, S.; SMITH, T.; SONSTEGARD, T.; VAN TASSEL, C.** 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PloS ONE.* 4: 1-13.
- MEUWISSEN, T.; HAYES, B.; GODDARD, M.** 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157(4): 1819-1829.
- MUIR, W.** 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124: 342-355.
- NIELSEN, H.; SONESSON, A.; YAZDI, H.; MEUWISSEN, T.** 2009. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture.* 289(3-4): 259-264.
- NIREA, K.; SONESSON A.; WOOLLIAMS, J.; MEUWISSEN, T.** 2012. Strategies for implementing genomic selection in family-based aquaculture breeding schemes: double haploid sib test populations. *Genet. Sel. Evol.* 44: 30-38.
- NORDBORG, M.; TAVARÉ, S.** 2002. Linkage disequilibrium: what story has to tell us. *Trends Genet.* 18(2): 83-90.
- NORDMO, R.; SEVATDAL, S.; RAMSTAD, A.** 1997. Experimental infection with *Vibrio salmonicida* in Atlantic salmon (*Salmo salar*): an evaluation of three different challenge methods. *Aquaculture.* 158: 23-32.
- ODEGARD, J.; BARANSKI, M.; GJERDE, B.; DJEDREM, T.** 2001. Methodology for genetic evaluation of disease resistance in aquaculture species: Challenges and future prospects. *Aquaculture.* 42: 103-114.
- ODEGARD, J.; MOEN, T.; SANTI, N.; KORSVOLL, S.; KJOGLUM, S.; MEUWISSEN, T.** 2014. Genomic prediction in an admixed population of Atlantic Salmon (*Salmo salar*). *Front. In Gen.* 5: 1-8.
- PALTI, Y.; GAO, G.; LIU, S.; KENT, M.; LIEN, S.; MILLER, M.; REXROAD, C.; MOENS, T.** 2015. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol. Ecol. Resour.* 15(3): 662-672.

- PÉREZ-ENCISO, M.; RINCÓN, J.; LEGARRA, A.** 2015. Sequence- vs. chip-assisted genomic selection: accurate biological information is advised. *Genet. Sel. Evol.* 47: 1-14.
- PÉREZ-ENCISO, M.; LEGARRA, A.** 2016. A combined coalescence gene-dropping tool for evaluating genomic selection in complex scenarios (ms2gs). *J. Anim. Breed. Genet.* 133(2): 85-91.
- PÉREZ-ENCISO, M.; FORNERIS, N.; DE LOS CAMPOS, G.; LEGARRA, A.** 2016. Evaluating sequence-based genomic prediction with an efficient new simulator. *Genetics*. DOI 10.1534/genetics.116.194878.
- PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.; BENDER, D.; MALLER, J.; SKLAR, P.; DE BAKKER, P.; DALY, M.; SHAM, P.** 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81(3): 559-575.
- RABERG, L.; SIM, D.; READ, A.** 2007. Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science.* 318: 812-814.
- RAMOS, A.; CROOIJMANS, A.; AFFARA, N.; AMARAL, A.; ARCHIBALD, A.; BEEVER, J.; BENDIXEN, C.; CHURCHER, C.; CLARK, R.; DEHAIS, P.; HANSEN, H.; HEDEGAARD, J.; HU, Z.; KERSTENS, H.; LAW, A.; MEGEN, H.; MILAN, D.; NONNEMAN, D.; ROHRER, G.; ROTHSCHILD, M.; SMITH, T.; SCHNABEL, R.; VAN TASSELL, C.; TAYLOR, J.; WIEDMANN, R.; SCHOOK, L.; GROENEN, A.** 2009. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. *PloS ONE.* 4: 1-13.
- ROBERTS, R.; PEARSON, M.** 2005. Infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *J. Fish. Dis.* 28: 383-390.
- ROZAS, M.; HENRÍQUEZ, R.** 2014. Piscirickettsiosis and *Piscirickettsia salmonis* in fish: A review. *J. Fish. Dis.* 37: 163-188.
- SARGOLZAEI, M.; SCHENKEL, F.** 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics.* 25(5): 680-681.

- SARGOLZAEI, M.; CHESNAIS, J.; SCHENKEL, F.** 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 15: 478-489.
- SERNAPESCA.** 2016. Informe sanitario de salmonicultura en centros marinos. Servicio Nacional de Pesca y Acuicultura. Departamento de Sanidad Pesquera. Chile. 37 p.
- SHIRALI, M.; MIRAEI-ASHTIANI, S.; PAKDEL, A.; HALEY, C.; NAVARRO, P.; PONG-WONG, R.** 2015. A comparison of the sensitivity of the Bayes C and Genomic Best Linear Unbiased Prediction (GBLUP) methods of estimating genomic breeding values under different Quantitative Trait Loci (QTL) model assumptions. *Ir. J. App. An. Sci.* 5(1): 41-46.
- SMITH, P.; LARENAS, J.; VERA, P.; CONTRERAS, J.; VENEGAS, C.; ROJAS, M.; GUAJARDO, A.** 2001. Principales enfermedades de los peces salmonídeos cultivados en Chile. *Monografías de medicina veterinaria*. 21: 3-19.
- SONESSON, A.; MEUWISSEN, T.** 2009. Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* 41(1): 37-44.
- SUBPESCA.** 2017. Informe sectorial de pesca y acuicultura. Subsecretaría de Pesca y Acuicultura. Departamento de Análisis Sectorial. Chile. 19 p.
- TAYLOR, J.** 2014. Implementation and accuracy of genomic selection. *Aquaculture*. 421-420: S8-S14.
- TSAI, H.; HAMILTON, A.; TINCH, A.; GUY, D.; BRON, J.; TAGGART, J.; GHARBI, K.; STEAR, M.; MATIKA, O.; PONG-WONG, R.; BISHOP, E.; HOUSTON, R.** 2016. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48: 47-57.
- TSAI, H.; MATIKA, O.; MCKINNON, E.; ANTOLÍN-SÁNCHEZ, R.; HAMILTON, A.; GUY, D.; TINCH, A.; GHARBI, K.; STEAR, M.; TAGGART, J.; BRON, J.; HICKEY, J.; HOUSTON, R.** 2017. Genotype Imputation to Improve the Cost-Efficiency of Genomic Selection in Farmed Atlantic Salmon. *G3*. 7: 1377-1383.

- VALLEJO, R.; LEEDS, T.; FRAGOMENI, B.; GAO, G.; HERNANDEZ, A.; MISZTAL, I.; WELCH, T.; WIENS, G.; PALT, Y.** 2016. Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: Insights on genotyping methods and genomic prediction models. *Front. Genet.* 7(96): 1-13.
- VALLEJO, R.; LEEDS, T.; GAO, G.; PARSONS, J.; MARTIN, K.; EVENHUIS, J.; FRAGOMENI, B.; WIENS, G.; PALT, Y.** 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49: 17-29.
- VELA-AVITUÁ, S.; MEUWISSEN, T.; LUAN, T.; ODEGARD, O.** 2015. Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. *Genet. Sel. Evol.* DOI 10.1186/s12711-014-0084-2.
- VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.** 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Gen. Sel. Evol.* 34: 275-305.
- VILLANUEVA, B.; FERNÁNDEZ, J.; GARCÍA-CORTÉS, L.; VARONA, L.; DAETWYLER, H.; TORO, M.** 2011. Accuracy of genome-wide evaluation for disease resistance in aquaculture breeding programs. *J. Anim. Sc.* 89(11): 3433-3442.
- VITEZICA, Z.; AGUILAR, I.; MISZTAL, I.; LEGARRA, A.** 2011. Bias in genomic predictions for populations under selection. *Genet. Res.* 93: 357-366.
- WIENTJES, Y.; CALUS, M.; GODDARD, M.; HAYES, B.** 2014. Effect of genetic architecture on accuracy of multi breed genomic prediction. **In:** Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Vancouver, Canada. August 17-22th, American Society of Animal Science (ASAS). pp. 74.
- WOLC, A.; STICKER, C.; ARANGO, J.; SETTAR, P.; FULTON, J.; O’SULLIVAN, N.; PREISINGER, R.; HABIER, D.; FERNANDO, R.; GARRICK, D.; LAMONT, S.; DEKKERS, J.** 2011. Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Gen. Sel. Evol.* 43(5): 1-9.

- YÁÑEZ, J.; MARTÍNEZ, V.** 2010. Genetic factors involved in resistance to infectious diseases in salmonids and their application in breeding programs. *Archivos de Medicina Veterinaria*. 42: 1-13.
- YÁÑEZ, J.; BANGERA, R.; LHORENTE, J.; OYARZÚN, M.; NEIRA, R.** 2013. Quantitative genetic variation of resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *Aquaculture*. 414-415: 155-159.
- YÁÑEZ, J.; HOUSTON, R.; NEWMAN, S.** 2014a. Genetics and genomics of disease resistance in salmonid species. *Front. Gen.* 5(11): 1-13.
- YÁÑEZ, J.; LHORENTE, J.; BASSINI, L.; OYARZÚN, M.; NEIRA, R.; NEWMAN, S.** 2014b. Genetic co-variation between resistance against both *Caligus rogercresseyi* and *Piscirickettsia salmonis*, and body weight in Atlantic salmon (*Salmo salar*). *Aquaculture*. 433: 295-298.
- YÁÑEZ, J.; NASWA, S.; LÓPEZ, M.; BASSINI, L.; CABREJOS, M.; GILBEY, J.; BERNATCHEZ, L.; NORRIS, A.; SOTO, C.; EISENHART, J.; SIMPSON, B.; NEIRA, R.; LHORENTE, J.; SCHNABLE, P.; NEWMAN, S.; MILEHAM, A.; DEEB, N.** 2014c. Development of a 200K SNP array for Atlantic Salmon: Exploiting Across Continents Genetic Variation. **In:** Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Vancouver, Canada. August 17-22th, American Society of Animal Science (ASAS). pp. 3.
- YÁÑEZ, J.; NEWMAN, S.; HOUSTON, R.** 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Front. Gen.* DOI: 10.3389/fgene.2015.00128.
- YÁÑEZ, J.; NASWA, S.; LÓPEZ, M.; BASSINI, L.; CORREA, K.; GILBEY, J.; BERNATCHEZ, L.; NORRIS, A.; NEIRA, R.; LHORENTE, J.; SCHNABLE, P.; NEWMAN, S.; MILEHAM, A.; DEEB, N.; DI GENOVA, A.; MAAS, A.** 2016. Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Mol. Ecol. Resources*. 16: 1002-1011.

YOSHIDA, G.; CARVALHEIRO, R.; LHORENTE, J.; CORREA, K.; FIGUEROA, R.; HOUSTON, R.; YÁÑEZ, J. 2017. Accuracy of genotype imputation in two-generation farmed Atlantic salmon population using high density and low density SNP panels, Aquaculture (submitted).

ZHANG, Z.; ERBE, M.; HE, J.; OBER, U.; GAO, N.; ZHANG, H.; SIMIANER, H.; LI, J. 2015. Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. G3. 5: 615-62.