

# RESSALVA

Atendendo solicitação do(a)  
autor(a), o texto completo desta tese  
será disponibilizado somente a partir  
de 04/11/2023.

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP  
CÂMPUS DE JABOTICABAL**

**QUANTITATIVE GENETICS AND GENOMIC TOOLS  
APPLIED TO AQUACULTURE BREEDING**

**Baltasar Fernandes Garcia Neto  
Zootecnista**

**2021**

**UNIVERSIDADE ESTADUAL PAULISTA - UNESP  
CÂMPUS DE JABOTICABAL**

**QUANTITATIVE GENETICS AND GENOMIC TOOLS  
APPLIED TO AQUACULTURE BREEDING**

**Baltasar Fernandes Garcia Neto**

**Orientador: Dr. Roberto Carneiro**

**Coorientadores: Dr. Hugo Horácio Montaldo**

**Dra. Laiza Helena de Souza lung**

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para obtenção do título de Doutor em Genética e Melhoramento Animal

216q Garcia Neto, Baltasar Fernandes  
Quantitative genetics and genomic tools applied to  
aquaculture breeding / Baltasar Fernandes Garcia Neto. --  
Jaboticabal, 2021  
102 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),  
Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal  
Orientador: Roberto Carneiro  
Coorientador: Hugo Horácio Montaldo, Laiza Helena de  
Souza Lung

1. Uniformity of weight. 2. SNP panel. 3. Genotype  
imputation. 4. Genomics. 5. Aquaculture. I. Título.

**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA TESE: QUANTITATIVE GENETICS AND GENOMIC TOOLS APPLIED TO AQUACULTURE BREEDING


**AUTOR: BALTASAR FERNANDES GARCIA NETO**


**ORIENTADOR: ROBERTO CARVALHEIRO**

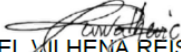
**COORDINADORA: LAIZA HELENA DE SOUZA IUNG**

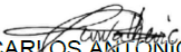
**COORDINADOR: HUGO HORÁCIO MONTALDO VALDENEGRO**


Aprovado como parte das exigências para obtenção do Título de Doutor em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:

  
Pesquisador Dr. ROBERTO CARVALHEIRO (Participação Virtual)  
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

  
Prof. Dr. DANISIO PRADO MUNARI (Participação Virtual)  
Departamento de Engenharia e Ciências Exatas (DECEX) / FCAV / Unesp - Jaboticabal

  
Prof. Dr. RAFAEL VILHENA REIS NETO (Participação Virtual)  
Curso de Engenharia de Pesca / UNESP, Câmpus de Registro, Registro-SP

  
Prof. Dr. CARLOS ANTONIO LOPES DE OLIVEIRA (Participação Virtual)  
Centro de Ciências Agrárias / Universidade Estadual de Maringá/PR

  
PhD. JOSÉ MANUEL YANEZ LOPEZ (Participação Virtual)  
Departamento de Medicina Veterinária Preventiva / Universidad del Chile (Santiago)

Jaboticabal, 04 de novembro de 2021

## **DADOS CURRICULARES DO AUTOR**

Baltasar Fernandes Garcia Neto nasceu em Jaboticabal, São Paulo, em 7 de outubro de 1991. Em 2010, iniciou o curso de graduação em Zootecnia na Universidade Estadual Paulista - UNESP em Jaboticabal. Desde o início do curso, Baltasar atuou na área de Aquicultura, mais especificamente na criação de camarões. Participou de diversos estudos no setor de Carcinicultura do CAUNESP e foi bolsista de pesquisa científica concedida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Em 2013, recebeu uma bolsa do programa Ciência sem Fronteiras concedida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para estudar por um ano na Irlanda, na National University of Ireland Galway - NUIG, onde conseguiu aprimorar suas habilidades em inglês e cursar algumas disciplinas. Em 2015, formou-se em Zootecnia pela UNESP, tendo como trabalho de conclusão de curso uma revisão bibliográfica sobre estimativas de herdabilidade em camarões, sob orientação do Dr. Roberto Carvalheiro. Em 2016, ele iniciou seu mestrado na UNESP, no programa de pós-graduação em genética e melhoramento animal, com bolsa da CAPES, sob orientação do Dr. Roberto Carvalheiro, terminando em 2018. O tema da dissertação foi a avaliação de diferentes métodos estatísticos para o mapeamento de QTLs usando simulações. Ainda em 2018, ingressou no curso de doutorado no mesmo programa de pós-graduação e sob a mesma orientação, com bolsa CAPES. Em 2020, foi contemplado com uma bolsa CAPES-PrInt que o permitiu realizar parte das atividades do doutorado na Universidade do Chile (Santiago, Chile) sob orientação do Dr. José Manuel Yáñez.

### **Dedico**

Aos meus pais, Baltasar e Silvana, à minha irmã, Carol, e à minha esposa, Rachel, pelo apoio e companheirismo durante essa jornada.

## AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por todas as bênçãos em minha vida.

Gostaria de agradecer ao Dr. Roberto Carvalheiro por todos os ensinamentos e conselhos. Agradeço por toda paciência, gentileza e por ser um exemplo de zootecnista: ético, profissional e acima de tudo humano.

Gostaria de agradecer a todos os membros das comissões examinadoras interna e externa: Dr. Danísio Prado Munari, Dr. Diogo Teruo Hashimoto, Dr. José Manuel Yáñez e Dr. Carlos Antônio Lopes de Oliveira pelo precioso tempo e atenção dispensados com esta tese e pelo seu valor inestimável nas sugestões e contribuições.

Agradeço à Dra. Laiza Helena de Souza lung por toda ajuda e paciência que além de coorientadora é uma excelente amiga.

Agradeço ao Dr. Hugo Horácio Montaldo, em nome da Universidad Nacional Autónoma de México (UNAM), que gentilmente nos forneceu os dados e contribuiu para o desenvolvimento do segundo capítulo desta tese.

Agradeço ao Dr. José Manuel Yáñez, em nome da Universidad de Chile (UChile), que me recebeu de forma muito acolhedora em seu laboratório, forneceu os dados para o desenvolvimento dos capítulos 3 e 4 da presente tese, e também contribuiu ativamente no desenvolvimento destes capítulos.

Agradeço as empresas: Maricultura del Pacífico (México), OPUMARSA (Equador), Aquacorporación Internacional (Costa Rica) e AquaAmerica (Brasil) pelo fornecimento dos dados utilizados nesta tese.

Agradeço ao programa de pós-graduação em Genética e Melhoramento Animal e o curso de graduação em Zootecnia da FCAV- UNESP Jaboticabal. Obrigado por todo conhecimento que me foi apresentado desde 2010. Agradeço à Dra. Sandra Aidar de Queiroz, Dra. Lucia Galvão de Albuquerque e Dr. Danísio Prado Munari, professores em minha graduação e pós-graduação. Obrigado por todo o conhecimento passado e incentivo ao estudo do melhoramento genético animal.

Agradeço aos meus amigos do departamento de Zootecnia e companheiros da “salinha”: Laiza, Thaise, Daiane, Giovana, André Mauric, Andrés (Colômbia), Lúcio, Caio (Índio), Diego (Pumba), Rafael (Japa), Willian, Ivan, Thales, Isabella, Delvan,



Kétuly e Gabriel, pela ajuda em assuntos acadêmicos e também pelos momentos de confraternização nos cafés e churrascos.

Agradeço aos integrantes do “Laboratorio de genómica aplicada a la Acuicultura” na FAVET (UChile): Paulina, Grazyella, Carolina, Lucas, Rodrigo, Pablo, Sebastián, Tamara, David e Jouseph, por me receberem tão bem no grupo de pesquisa e me fazerem sentir em “casa” durante os 12 meses que estive no Chile.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## SUMMARY

RESUMO .....	iv
ABSTRACT .....	vi
CHAPTER 1 – General Considerations .....	1
1.1 Introduction .....	1
1.2 Literature Review .....	2
1.2.1 Uniformity .....	2
1.2.1.1 Genetics of uniformity in aquaculture species .....	2
1.2.1.2 Models to evaluate the uniformity .....	3
1.2.1.3 Estimation of uniformity genetic parameters .....	5
1.2.2 Genomic tools applied to aquaculture breeding .....	6
1.2.2.1 Genomic resources and genotyping methods in aquaculture .....	6
1.2.2.2 Linkage disequilibrium .....	7
1.2.2.3 Genotype imputation .....	9
1.3 Objectives .....	10
1.4 References .....	11
CHAPTER 2 – Effect of harvest weight and its uniformity on survival in <i>Litopenaeus vannamei</i> reared in different systems .....	20
ABSTRACT .....	20
2.1 Introduction .....	21
2.2 Material and methods .....	23
2.2.1 Population background .....	23
2.2.2 Larvae and post-larvae management .....	23
2.2.3 Collection and edition of phenotypes .....	24
2.2.4 Data structure and statistical analysis .....	24
2.2.5 Harvest weight and its uniformity .....	26
2.2.6 Harvest weight and survival .....	27
2.2.7 Genetic parameter estimations .....	28
2.3. Results .....	29
2.3.1 Descriptive statistics .....	29
2.3.2 Genetic parameters for harvest weight, uniformity and survival .....	29
2.3.3 Genetic correlations among traits .....	30
2.4. Discussion .....	32
2.4.1 Genetic parameters for harvest weight and uniformity .....	32
2.4.2 Genetic parameters for harvest weight and survival .....	33

2.4.3 Genetic correlations .....	34
2.5 Conclusion .....	36
2.6 References.....	37
CHAPTER 3 – Application of a novel 50K SNP genotyping array to assess the genetic diversity and linkage disequilibrium in a farmed Pacific white shrimp ( <i>Litopenaeus vannamei</i> ) population .....	45
ABSTRACT.....	45
3.1 Introduction .....	46
3.2 Material and methods.....	48
3.2.1 Population and samples.....	48
3.2.2 Genotyping, quality control and genetic diversity .....	48
3.2.3 LD estimation .....	49
3.2.4 Effective population size ( $N_e$ ) .....	50
3.3 Results.....	51
3.3.1 SNP array validation and quality control .....	51
3.3.2 Genetic diversity .....	52
3.3.3 LD and $N_e$ estimation .....	53
3.4 Discussion.....	57
3.4.1 Genetic diversity .....	57
3.4.2 LD and $N_e$ .....	58
3.4.3 Practical applications.....	61
3.5 Conclusion .....	62
3.6 References.....	63
CHAPTER 4 – Accuracy of genotype imputation to sequence level using different populations of Nile tilapia .....	73
ABSTRACT .....	73
4.1 Introduction .....	75
4.2 Material and methods.....	76
4.2.1 Origin of animals .....	76
4.2.2 Whole-genome sequencing and quality control of genotypes .....	77
4.2.3 Imputation scenarios .....	78
4.2.4 Genotype imputation and accuracy.....	79
4.3 Results.....	80
4.3.1 WGS and quality control .....	80
4.3.2 Accuracy of genotype imputation .....	81
4.3.2.1 Effect of reference population size and origin .....	82
4.3.2.2 Effect of MAF .....	83
4.3.2.3 Effect of sequencing coverage.....	83

4.4 Discussion.....	85
4.5 Conclusion .....	90
4.6 References.....	91
CHAPTER 5 – Final Considerations.....	98
APPENDICES.....	100
Appendix A .....	101
Appendix B .....	102
Appendix C .....	102

## GENÉTICA QUANTITATIVA E FERRAMENTAS GENÔMICAS APLICADAS AO MELHORAMENTO GENÉTICO NA AQUICULTURA

**RESUMO** – A procura por animais mais uniformes frente as adversidades ambientais e a aplicação de ferramentas genômicas são exemplos de novas estratégias para tornar o melhoramento genético em espécies aquícolas mais eficiente. No intuito de avaliar a efetividade destas estratégias, os objetivos do presente trabalho foram: i) estimar os componentes genéticos da uniformidade de peso a despesca (PD) e investigar se o PD e sua uniformidade ( $PD_u$ ) podem afetar a sobrevivência (SOB) em camarões; ii) estimar o desequilíbrio de ligação (LD) em uma população de camarões cultivados usando um novo painel de 50k polimorfismos de nucleotídeos únicos (SNPs) e iii) avaliar a viabilidade da imputação ao nível de sequência em tilápia do Nilo e os impactos do tamanho e origem da população de referência na acurácia. Para i), foram utilizados 149.919 registros de PD e 164.023 de SOB de uma larvicultura de camarões mexicana. Os dados foram agrupados em três conjuntos de acordo com o tipo de produção: viveiros escavados com baixa densidade (S1), tanques de concreto com recirculação em alta densidade (S2) e juntando ambos conjuntos de dados (S1+S2). Um modelo linear generalizado hierárquico duplo foi aplicado para estimar os componentes de (co)variância de PD- $PD_u$  e um modelo bivariado linear misto foi utilizado para estimar os componentes referentes a PD-SOB. As correlações genéticas ( $r_g$ ) para estas características foram estimadas, além da correlação entre valores genéticos (VG) para  $PD_u$ -SOB. Para ii), 96 camarões (40 machos e 56 fêmeas) reprodutores originários de uma larvicultura do Equador foram genotipados utilizando um novo painel de 50k SNPs. Posteriormente, o LD foi estimado usando três controles de qualidade distintos com diferentes filtros de frequência de alelo menor: 0,1 (CQ1); 0,05 (CQ2) e 0,01 (CQ3). Para iii), amostras de DNA de 326 tilápias provenientes de três populações de origens distintas ( $P_A$ ,  $P_B$  e  $P_C$ ) foram extraídas e sequenciadas. Após o controle de qualidade dos dados de sequência foram obtidos 4,6 milhões de SNPs em comum para todas as populações. A imputação foi feita em quatro cenários distintos: dois tamanhos (10 ou 90% dos animais de cada população) e duas origens de referência (apenas duas populações diferentes ou todas as três populações). Os animais de validação tiveram seus genótipos ocultados mantendo somente 50k SNPs para a imputação a nível de sequência. Foi utilizado o software FImpute3 e a acurácia de imputação foi avaliada através da correlação entre o genótipo imputado e observado ( $r^2$ ). No estudo de uniformidade, uma proporção significativa de variância genética foi detectada na variância residual de PD, revelando que existe a possibilidade de selecionar para uniformidade desta característica em camarões (coeficiente de variação genética variando de 17 a 35%). As  $r_g$  entre PD e SOB foram diferentes em sinal e magnitude com base no sistema de produção, sendo iguais a 0,36, -0,59 e -0,02 para S1, S2 e S1+S2, respectivamente. Os coeficientes de correlação entre o VG de  $PD_u$  e SOB foram significativos apenas para S1 (-0,32), sugerindo que a seleção para famílias mais uniformes também pode aumentar as taxas de sobrevivência em um ambiente de menor densidade. Para os dados genômicos de camarões, foram obtidos 34.425, 39.091 e 42.789 SNPs após CQ1, CQ2 e CQ3, respectivamente, exibindo alto grau de polimorfismo e validando a aplicação deste painel de SNPs para esta população de camarões cultivados. O LD decaiu rapidamente nos primeiros 30 KB de distância de 0,2 para 0,07 e depois diminuiu para 0,02 para distancias mais longas (>80 KB). Esses resultados sugerem

incorporação recente de animais de diferentes populações ou linhagens neste grupo de reprodutores e que a seleção genômica e estudos de associação amplo do genoma são viáveis em camarões usando este painel de SNP como ferramenta. Para a imputação, no geral, o  $r^2$  por animal mostrou resultados intermediários variando de 0,37 a 0,56 para  $P_A$  e 0,43 a 0,58 para  $P_B$  e  $P_C$ . Entretanto, os resultados mostraram que foi possível imputar de 50k a aproximadamente 680k com alta acurácia usando dados de sequência de tilápia. Foi observado um aumento de 31,5% para  $P_A$  e 24,6% para  $P_B$  e  $P_C$  no  $r^2$ , quando 90% dos animais da mesma população foram usados como referência em comparação a 10%. Não houve diferenças significativas para  $r^2$  entre os cenários que usaram 90% dos animais da mesma população e usaram animais das três populações como referência, mostrando que a estratégia de usar informações de outra população para aumentar a população de referência teve pouco efeito na acurácia de imputação.

Palavras-chave: Camarão, Desequilíbrio de ligação, Imputação de genótipos, Tilápia, Uniformidade

## QUANTITATIVE GENETICS AND GENOMIC TOOLS APPLIED TO AQUACULTURE BREEDING

**ABSTRACT** – The search for more uniform animals in the face of environmental adversities and the application of genomic tools are examples of new strategies to make genetic improvement in aquaculture species more efficient. In order to evaluate the effectivity of these strategies, the aims of the present study were: i) to estimate the genetic components of uniformity of harvest weight (HW) and investigate whether the HW and its uniformity ( $HW_v$ ) may affect the survival (SUR) in shrimp; ii) to estimate linkage disequilibrium (LD) in a population of farmed shrimp using a new 50k single-nucleotide polymorphisms (SNPs) panel and; iii) to assess the feasibility of genotype imputation to sequence level in Nile tilapia and the impacts of size and origin of population reference on accuracy of imputation. For i), 149,919 records of HW and 164,023 of SUR were obtained from a Mexican shrimp hatchery. The data were grouped into three sets according to the system of production: excavated ponds with low density (S1), concrete recirculation tanks with high density (S2) and joining both sets of data (S1+S2). A double hierarchical generalized linear model was applied to estimate the (co)variance components of HW- $HW_v$  and a bivariate mixed linear model was used to estimate the components referring to HW-SUR. The genetic correlations ( $r_g$ ) for these traits were estimated, in addition to the correlation between the estimated breeding values (EBV) for  $HW_v$ -SUR. For ii), 96 broodstock shrimp (40 males and 56 females) from a commercial hatchery from Ecuador were genotyped using a new 50k SNPs panel. Then, the LD was estimated using three distinct quality controls with different minor allele frequency filters: 0.1 (QC1); 0.05 (QC2) and 0.01 (QC3). For iii), DNA samples from 326 tilapia from three populations ( $P_A$ ,  $P_B$  and  $P_C$ ) were extracted and sequenced. After quality control of sequence data, 4.6 million of SNPs were obtained in common for all populations. The imputation was performed in four different scenarios: two sizes (10 or 90% of animals from each population) and two origins of population reference (only two different populations or all three populations). The validation animals had their genotypes masked keeping only 50k SNPs for imputation to the sequence level. The FImpute3 software was used and the imputation accuracy was evaluated through the correlation between the imputed and observed genotypes ( $r^2$ ). In the uniformity study, a significant proportion of genetic variance was detected in the residual variance of HW, revealing that there is possibility to select for uniformity of this trait in shrimp (genetic coefficient of variation ranging between 17 and 35%). The genetic correlations between HW and SUR were different in sign and magnitude based on the production system, being equal to 0.36, -0.59 and -0.02 for S1, S2 and S1+S2, respectively. The correlation coefficients between the EBV of  $HW_v$  and SUR were significant only for S1 (-0.32), suggesting that selection for more uniform families may also increase survival rates in a lower density environment. For shrimp genomic data, 34,425, 39,091 and 42,789 SNPs were obtained after QC1, QC2 and QC3, respectively, showing high polymorphism and validating the application of this panel of SNPs to this farmed shrimp population. The LD declined rapidly in the first 30 KB of distance from 0.2 to 0.07 and then decreased to 0.02 for longer distances (>80 KB). These results suggest recent incorporation of animals from different populations or strains into this breeding population and that genomic selection and genome-wide association studies are feasible in shrimp using this SNP panel as tool. For the imputation using tilapia sequence data, in general, the  $r^2$  per animal showed intermediate results ranging from 0.37 to 0.56 for  $P_A$  and 0.43 to 0.58 for  $P_B$  and  $P_C$ .

However, the results showed that it was possible to impute from 50k to approximately 680k with high accuracy. An increase in  $r^2$  of 31.5% for  $P_A$  and 24.6% for  $P_B$  and  $P_C$  was observed, when 90% of animals from the same population were used as reference. There were no significant differences for  $r^2$  between the scenarios that used 90% of animals from the same population and used animals from the three populations as a reference, showing that the strategy of using information from another population to increase the reference population had minor effect on the accuracy of imputation.

Keywords: Shrimp, Linkage Disequilibrium, Genotype Imputation, Tilapia, Uniformity



## **CHAPTER 1 – General Considerations**

### **1.1 Introduction**

Currently, aquaculture has played an important role in food production. In 2019, aquaculture was responsible by the production of approximately 120.1 million of tons representing an increase of 54% in comparison to 2009. Similar tendency was observed in the Brazilian sector, which approximately 600.3 thousands of tons were produced in 2019 (45.7% more than the production registered ten years ago) (FAO, 2021). Shrimp and tilapia farming were segments that represented large proportions of the total aquaculture production in Brazil in 2019, 9.1 and 53.9%, respectively (FAO, 2021).

The implementation of breeding programs is a strategy employed to improve the mean level of traits that present economic return. Traits related to growth (e.g., weight at harvest or specific age), survival (e.g., survival in specific production systems and after disease challenge), reproduction (e.g., age at sexual maturation and number of eggs) and meat quality (e.g., fillet color and yield) are groups of traits often contemplated in the aquaculture breeding programs (Gjedrem and Baranski, 2009). Although breeding programs focused on improving the mean level of these traits have been successful (Gjedrem and Rye, 2018), there is a new concern to decrease the phenotypic variability of such traits, i.e., to increase uniformity. In order to select for more uniform animals, the low residual (environmental) variance may be used as selection criterium, if this is a trait under genetic control (Mulder et al., 2008). Despite its importance to aquaculture industry and relevance to the effectivity of breeding programs, uniformity studies are scarce in this field.

Another strategy to increase the effectivity of breeding programs is the inclusion of genomic information. With the advent of new technologies, it is possible to extract DNA from animals, genotype this samples and include this as additional source of information. For breeding programs, the genomic information has two main applications: genomic selection (GS) and genome-wide association studies (GWAS). GS consists of estimating the genomic breeding values of selection candidates using a prediction equation that was previously calculated in a reference population (Meuwissen et al., 2001). GS requires a sufficient number (thousands) of single-nucleotide polymorphisms (SNPs) markers, allowing to capture genetic variability

through linkage disequilibrium (LD) with quantitative trait loci (QTL). On the other hand, GWAS may help to disentangle the genetic architecture of traits and detect genomic regions associated to genetic variation by using LD information between SNPs and QTLs. In this sense, when a SNP panel is developed, it is important to determine the level and extension of LD to evaluate its applicability in breeding programs.

The use of whole-genome sequencing (WGS) to obtain genotypic information is expected to accelerate the genetic gain in comparison to genotyping panels because it may include millions of SNPs with potential causal variants related to relevant phenotypes (Meuwissen and Goddard, 2010). Nevertheless, WGS faces an important drawback compared to SNP panels: it is much more expensive mainly when many animals are required. In this sense, the genotype imputation is a cost-effective strategy to obtain genotypes at sequence level. The imputation is performed using WGS information from key-animals as reference to predict the missing genotypes of target animals that were previously genotyped in a lower density. The size and origin of reference population are important aspects that should be evaluated to perform genotype imputation with high accuracy (Carvalho et al., 2014; Hickey et al., 2012). In spite of the increasing number of WGS studies in aquaculture species, none study was found evaluating how these factors may affect the final accuracy of imputation.

#### **4.5 Conclusion**

It was possible to impute from 50k to approximately 680k SNPs using tilapia WGS data with expected accuracy between 0.92 and 0.95, considering only SNPs imputed with accuracy greater than 0.8. The use of more animals from same population as reference had higher impact in the accuracy of imputation than animals from different populations. However, a higher proportion of SNPs imputed with high accuracy was obtained when animals from all populations were used as reference.

#### 4.6 References

- Binsbergen, R. van, Bink, M.C., Calus, M.P., Eeuwijk, F.A. van, Hayes, B.J., Hulsegge, I., Veerkamp, R.F., 2014. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. *Genet. Sel. Evol.* 2014 461 46, 1–13. <https://doi.org/10.1186/1297-9686-46-41>
- Bolormaa, S., Chamberlain, A.J., Khansefid, M., Stothard, P., Swan, A.A., Mason, B., Prowse-Wilkins, C.P., Duijvesteijn, N., Moghaddar, N., Werf, J.H. van der, Daetwyler, H.D., MacLeod, I.M., 2019. Accuracy of imputation to whole-genome sequence in sheep. *Genet. Sel. Evol.* 2019 511 51, 1–17. <https://doi.org/10.1186/S12711-018-0443-5>
- Bouwman, A.C., Veerkamp, R.F., 2014. Consequences of splitting whole-genome sequencing effort over multiple breeds on imputation accuracy. *BMC Genet.* 2014 151 15, 1–9. <https://doi.org/10.1186/S12863-014-0105-8>
- Browning, B.L., Browning, S.R., 2008. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84, 210–223. <https://doi.org/10.1016/j.ajhg.2009.01.005>
- Cáceres, G., López, M.E., Cádiz, M.I., Yoshida, G.M., Jedlicki, A., Palma-Véjares, R., Travisany, D., Díaz-Domínguez, D., Maass, A., Lhorente, J.P., Soto, J., Salas, D., Yáñez, J.M., 2019. Fine Mapping Using Whole-Genome Sequencing Confirms Anti-Müllerian Hormone as a Major Gene for Sex Determination in Farmed Nile Tilapia (*Oreochromis niloticus* L.). *G3 Genes, Genomes, Genet.* 9, 3213–3223. <https://doi.org/10.1534/g3.119.400297>
- Calus, M.P.L., Bouwman, A.C., Hickey, J.M., Veerkamp, R.F., Mulder, H.A., 2014. Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications. *animal* 8, 1743–1753. <https://doi.org/10.1017/S1751731114001803>
- Carvalho, R., Boison, S.A., Neves, H.H.R., Sargolzaei, M., Schenkel, F.S., Utsunomiya, Y.T., O'Brien, A.M.P., Sölkner, J., McEwan, J.C., Van Tassell, C.P., Sonstegard, T.S., Garcia, J.F., 2014. Accuracy of genotype imputation in Nelore cattle. *Genet. Sel. Evol.* 46, 1–11. <https://doi.org/10.1186/s12711-014-0069-1>
- Conte, M.A., Gammerdinger, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D.,

2017. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* 18, 1–19. <https://doi.org/10.1186/s12864-017-3723-5>

Conte, M.A., Joshi, R., Moore, E.C., Nandamuri, S.P., Gammerdinger, W.J., Roberts, R.B., Carleton, K.L., Lien, S., Kocher, T.D., 2019. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *Gigascience* 8, 1–20. <https://doi.org/10.1093/gigascience/giz030>

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P.J., Coote, D., Chamberlain, A.J., Anderson, C., Vantassell, C.P., Hulsegge, I., Goddard, M.E., Guldbbrandtsen, B., Lund, M.S., Veerkamp, R.F., Boichard, D.A., Fries, R., Hayes, B.J., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46, 858–865. <https://doi.org/10.1038/ng.3034>

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Group, 1000 Genomes Project Analysis, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/BIOINFORMATICS/BTR330>

Druet, T., Macleod, I.M., Hayes, B.J., 2013. Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Hered.* 2014 1121 112, 39–47. <https://doi.org/10.1038/hdy.2013.13>

Dufflocq, P., Pérez-Enciso, M., Lhorente, J.P., Yáñez, J.M., 2019. Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: A simulation study. *Aquaculture* 503, 225–230. <https://doi.org/10.1016/j.aquaculture.2018.12.061>

Eknath, A.E., Hulata, G., 2009. Use and exchange of genetic resources of Nile tilapia (*Oreochromis niloticus*). *Rev. Aquac.* 1, 197–213. <https://doi.org/10.1111/j.1753-5131.2009.01017.x>

Eknath, A.E., Tayamen, M.M., Palada-de Vera, M.S., Danting, J.C., Reyes, R.A., Dionisio, E.E., Capili, J.B., Bolivar, H.L., Abella, T.A., Circa, A. V., Bentsen, H.B.,

- Gjerde, B., Gjedrem, T., Pullin, R.S.V., 1993. Genetic improvement of farmed tilapias: the growth performance of eight strains of *Oreochromis niloticus* tested in different farm environments. *Aquaculture* 111, 171–188. [https://doi.org/10.1016/0044-8486\(93\)90035-W](https://doi.org/10.1016/0044-8486(93)90035-W)
- FAO, 2021. FAO Fisheries & Aquaculture - Statistics [WWW Document]. URL <http://www.fao.org/fishery/statistics/en> (accessed 6.30.21).
- Fernandes Júnior, G.A., Carvalheiro, R., Oliveira, H.N., Sargolzaei, M., Costilla, R., Ventura, R. V., Fonseca, L.F.S., Neves, H.H.R., Hayes, B.J., de Albuquerque, L.G., 2021. Imputation accuracy to whole-genome sequence in Nellore cattle. *Genet. Sel. Evol.* 53, 27. <https://doi.org/10.1186/s12711-021-00622-5>
- Goddard, M.E., Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. <https://doi.org/10.1038/nrg2575>
- Hayes, B.J., Bowman, P.J., Daetwyler, H.D., Kijas, J.W., Van Der Werf, J.H.J., 2012. Accuracy of genotype imputation in sheep breeds. *Anim. Genet.* 43, 72–80. <https://doi.org/10.1111/j.1365-2052.2011.02208.x>
- Hickey, J.M., Crossa, J., Babu, R., de los Campos, G., 2012a. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Sci.* 52, 654–663. <https://doi.org/10.2135/cropsci2011.07.0358>
- Hickey, J.M., Kinghorn, B.P., Tier, B., van der Werf, J.H., Cleveland, M.A., 2012b. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genet. Sel. Evol.* 2012 441 44, 1–11. <https://doi.org/10.1186/1297-9686-44-9>
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., Abecasis, G.R., 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 2012 448 44, 955–959. <https://doi.org/10.1038/ng.2354>
- Huang, L., Li, Y., Singleton, A.B., Hardy, J.A., Abecasis, G., Rosenberg, N.A., Scheet, P., 2009. Genotype-Imputation Accuracy across Worldwide Human Populations. *Am. J. Hum. Genet.* 84, 235–250. <https://doi.org/10.1016/J.AJHG.2009.01.013>
- Joshi, R., Almeida, D.B., da Costa, A.R., Skaarud, A., de Pádua Pereira, U., Knutsen, T.M., Moen, T., Alvarez, A.T., 2021a. Genomic selection for resistance

to Francisellosis in commercial Nile tilapia population: Genetic and genomic parameters, correlation with growth rate and predictive ability. *Aquaculture* 537, 736515. <https://doi.org/10.1016/j.aquaculture.2021.736515>

Joshi, R., Skaarud, A., Alvarez, A.T., Moen, T., Ødegård, J., 2021b. Bayesian genomic models boost prediction accuracy for survival to *Streptococcus agalactiae* infection in Nile tilapia (*Oreochromis niloticus*). *Genet. Sel. Evol.* 53, 37. <https://doi.org/10.1186/s12711-021-00629-y>

Joshi, R., Skaarud, A., de Vera, M., Alvarez, A.T., Ødegård, J., 2020. Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 516, 734641. <https://doi.org/10.1016/j.aquaculture.2019.734641>

Liu, Q., Cirulli, E.T., Han, Y., Yao, S., Liu, S., Zhu, Q., 2015. Systematic assessment of imputation performance using the 1000 Genomes reference panels. *Brief. Bioinform.* 16, 549–562. <https://doi.org/10.1093/BIB/BBU035>

Lu, S., Zhu, J., Du, X., Sun, S., Meng, L., Liu, S., Fan, G., Wang, J., Chen, S., 2020. Genomic selection for resistance to *Streptococcus agalactiae* in GIFT strain of *Oreochromis niloticus* by GBLUP, wGBLUP, and BayesCπ. *Aquaculture* 523, 735212. <https://doi.org/10.1016/j.aquaculture.2020.735212>

Ma, P., Brøndum, R.F., Zhang, Q., Lund, M.S., Su, G., 2013. Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. *J. Dairy Sci.* 96, 4666–4677. <https://doi.org/10.3168/JDS.2012-6316>

MacLeod, I.M., Bowman, P.J., Vander Jagt, C.J., Haile-Mariam, M., Kemper, K.E., Chamberlain, A.J., Schrooten, C., Hayes, B.J., Goddard, M.E., 2016. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* 17, 144. <https://doi.org/10.1186/s12864-016-2443-6>

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of Total

Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* 157, 1819–1829.

Neira, R., 2010. Breeding in Aquaculture Species: Genetic Improvement Programs in Developing Countries, in: 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany, p. 8.

Pausch, H., MacLeod, I.M., Fries, R., Emmerling, R., Bowman, P.J., Daetwyler, H.D., Goddard, M.E., 2017. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet. Sel. Evol.* 2017 49, 1–14. <https://doi.org/10.1186/S12711-017-0301-X>

Pérez-Enciso, M., Rincón, J.C., Legarra, A., 2015. Sequence- vs. chip-assisted genomic selection: Accurate biological information is advised. *Genet. Sel. Evol.* 47, 1–14. <https://doi.org/10.1186/s12711-015-0117-5>

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. <https://doi.org/10.1086/519795>

Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., Delaneau, O., 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* 2021 53, 120–126. <https://doi.org/10.1038/s41588-020-00756-0>

Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15, 1–12. <https://doi.org/10.1186/1471-2164-15-478>

Sukhavachana, S., Tongyoo, P., Massault, C., McMillan, N., Leungnaruemitchai, A., Poompuang, S., 2020. Genome-wide association study and genomic prediction for resistance against *Streptococcus agalactiae* in hybrid red tilapia (*Oreochromis* spp.). *Aquaculture* 525, 735297. <https://doi.org/10.1016/j.aquaculture.2020.735297>

Tsai, H.Y., Matika, O., Edwards, S.M.K., Antolín-Sánchez, R., Hamilton, A., Guy, D.R., Tinch, A.E., Gharbi, K., Stear, M.J., Taggart, J.B., Bron, J.E., Hickey, J.M., Houston, R.D., 2017. Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3 Genes, Genomes, Genet.* 7, 1377–1383. <https://doi.org/10.1534/g3.117.040717>



- Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E., Houston, R.D., 2020. Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. *G3 Genes, Genomes, Genet.* 10, 581–590. <https://doi.org/10.1534/g3.119.400800>
- VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. <https://doi.org/10.3168/JDS.2007-0980>
- VanRaden, P.M., Sun, C., O'Connell, J.R., 2015. Fast imputation using medium or low-coverage sequence data. *BMC Genet.* 2015 161 16, 1–12. <https://doi.org/10.1186/S12863-015-0243-7>
- Wood, S.N., 2006. On confidence intervals for generalized additive models based on penalized regression splines. *Aust. N. Z. J. Stat.* 48, 445–464. <https://doi.org/10.1111/J.1467-842X.2006.00450.X>
- Yáñez, J.M., Joshi, R., Yoshida, G.M., 2020a. Genomics to accelerate genetic improvement in tilapia. *Anim. Genet.* 51, 658–674. <https://doi.org/10.1111/AGE.12989>
- Yáñez, J.M., Newman, S., Houston, R.D., 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Front. Genet.* 6, 1–3. <https://doi.org/10.3389/fgene.2015.00128>
- Yáñez, J.M., Yoshida, G., Barria, A., Palma-Véjares, R., Travisany, D., Díaz, D., Cáceres, G., Cádiz, M.I., López, M.E., Lhorente, J.P., Jedlicki, A., Soto, J., Salas, D., Maass, A., 2020b. High-Throughput Single Nucleotide Polymorphism (SNP) Discovery and Validation Through Whole-Genome Resequencing in Nile Tilapia (*Oreochromis niloticus*). *Mar. Biotechnol.* 22, 109–117. <https://doi.org/10.1007/s10126-019-09935-5>
- Yoshida, G.M., Barria, A., Correa, K., Cáceres, G., Jedlicki, A., Cadiz, M.I., Lhorente, J.P., Yáñez, J.M., 2019a. Genome-Wide Patterns of Population Structure and Linkage Disequilibrium in Farmed Nile Tilapia (*Oreochromis niloticus*). *Front. Genet.* 10, 745. <https://doi.org/10.3389/fgene.2019.00745>
- Yoshida, G.M., Carvalheiro, R., Lhorente, J.P., Correa, K., Figueroa, R., Houston, R.D., Yáñez, J.M., 2018. Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture* 491, 147–154. <https://doi.org/10.1016/j.aquaculture.2018.03.004>

- Yoshida, G.M., Lhorente, J.P., Correa, K., Soto, J., Salas, D., Yáñez, J.M., 2019b. Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3 Genes, Genomes, Genet.* 9, 2597–2607. <https://doi.org/10.1534/g3.119.400116>
- Yoshida, G.M., Yáñez, J.M., 2021a. Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics* 22, 1–13. <https://doi.org/10.1186/s12864-020-07341-z>
- Yoshida, G.M., Yáñez, J.M., 2021b. Increased accuracy of genomic predictions for growth under chronic thermal stress in rainbow trout by prioritizing variants from GWAS using imputed sequence data. *Evol. Appl.* 00, 1–16. <https://doi.org/10.1111/eva.13240>
- Yu, X., Megens, H.-J., Mengistu, S.B., Bastiaansen, J.W.M., Mulder, H.A., Benzie, J.A.H., Groenen, M.A.M., Komen, H., 2021. Genome-wide association analysis of adaptation to oxygen stress in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* 22, 426. <https://doi.org/10.1186/s12864-021-07486-5>
- Zhang, Z., Druet, T., 2010. Marker imputation with low-density marker panels in Dutch Holstein cattle. *J. Dairy Sci.* 93, 5487–5494. <https://doi.org/10.3168/JDS.2010-3501>