



ANIMAL GENETICS AND DISEASES: ADVANCES IN FARMING AND LIVESTOCK SYSTEMS

EDITED BY: Mark S. Fife, John A. Hammond and Andrea B. Doeschl-Wilson
PUBLISHED IN: *Frontiers in Genetics*



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-523-8

DOI 10.3389/978-2-88963-523-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ANIMAL GENETICS AND DISEASES: ADVANCES IN FARMING AND LIVESTOCK SYSTEMS

Topic Editors:

Mark S. Fife, AVIAGEN UK Ltd, United Kingdom

John A. Hammond, Pirbright Institute, United Kingdom

Andrea B. Doeschl-Wilson, University of Edinburgh, United Kingdom

The new Animal Genetics and Disease 2017 conference committee organized a Research Topic for the proceedings of this inaugural conference. The meeting brought together specialists working on the interface between genomics, genetic engineering, and infectious disease, with the aims of improving animal and human health and welfare. This conference was funded by Advanced Courses and Scientific Conference at the Wellcome Genome Campus, Hinxton, UK.

The conference will highlight breakthroughs in genomic technologies that are rapidly increasing our understanding of the fundamental role that host and pathogen genetics play in infections and epidemics. This Research Topic focuses on how infections spread and how they further affect the productivity of livestock systems and food supply chains. Thanks to technological advances, we now have the tools for real-time surveillance of zoonoses affecting wildlife, farm animals and animal-to-human disease transmission.

Citation: Fife, M. S., Hammond, J. A., Doeschl-Wilson, A. B., eds. (2020). Animal Genetics and Diseases: Advances in Farming and Livestock Systems. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-523-8

Table of Contents

- 05** *Transcriptional Innate Immune Response of the Developing Chicken Embryo to Newcastle Disease Virus Infection*
Megan A. Schilling, Robab Katani, Sahar Memari, Meredith Cavanaugh, Joram Buza, Jessica Radzio-Basu, Fulgence N. Mpenda, Melissa S. Deist, Susan J. Lamont and Vivek Kapur
- 14** *Gene Expression Response to Sea Lice in Atlantic Salmon Skin: RNA Sequencing Comparison Between Resistant and Susceptible Animals*
Diego Robledo, Alejandro P. Gutiérrez, Agustín Barría, José M. Yáñez and Ross D. Houston
- 24** *RNA Sequencing (RNA-Seq) Reveals Extremely Low Levels of Reticulocyte-Derived Globin Gene Transcripts in Peripheral Blood From Horses (*Equus caballus*) and Cattle (*Bos taurus*)*
Carolina N. Correia, Kirsten E. McLoughlin, Nicolas C. Nalpas, David A. Magee, John A. Browne, Kevin Rue-Albrecht, Stephen V. Gordon and David E. MacHugh
- 35** *Genetic Analysis of a Commercial Egg Laying Line Challenged With Newcastle Disease Virus*
Kaylee Rowland, Anna Wolc, Rodrigo A. Gallardo, Terra Kelly, Huaijun Zhou, Jack C. M. Dekkers and Susan J. Lamont
- 47** *Overexpression of Chicken IRF7 Increased Viral Replication and Programmed Cell Death to the Avian Influenza Virus Infection Through TGF-Beta/FoxO Signaling Axis in DF-1*
Tae Hyun Kim and Huaijun Zhou
- 59** *Applications of Gene Editing in Chickens: A New Era is on the Horizon*
Hicham Sid and Benjamin Schusser
- 71** *Including Phenotypic Causal Networks in Genome-Wide Association Studies Using Mixed Effects Structural Equation Models*
Mehdi Momen, Ahmad Ayatollahi Mehrgardi, Mahmoud Amiri Roudbar, Andreas Kranis, Renan Mercuri Pinto, Bruno D. Valente, Gota Morota, Guilherme J. M. Rosa and Daniel Gianola
- 82** *Genetic Parameters for Resistance to Non-specific Diseases and Production Traits Measured in Challenging and Selection Environments; Application to a Rabbit Case*
Mélanie Gunia, Ingrid David, Jacques Hurtaud, Mickaël Maupin, Hélène Gilbert and Hervé Garreau
- 92** *The Genomic Architecture of Fowl Typhoid Resistance in Commercial Layers*
Androniki Psifidi, Kay M. Russell, Oswald Matika, Enrique Sánchez-Molano, Paul Wigley, Janet E. Fulton, Mark P. Stevens and Mark S. Fife
- 103** *Sequence Characterization of DSG3 Gene to Know its Role in High-Altitude Hypoxia Adaptation in the Chinese Cashmere Goat*
Chandar Kumar, Shen Song, Lin Jiang, Xiaohong He, Qianjun Zhao, Yabin Pu, Kanwar Kumar Malhi, Asghar Ali Kamboh and Yuehui Ma

- 117** *Dissecting the Genomic Architecture of Resistance to Eimeria maxima Parasitism in the Chicken*
 Kay Boulton, Matthew J. Nolan, Zhiguang Wu, Valentina Riggio, Oswald Matika, Kimberley Harman, Paul M. Hocking, Nat Bumstead, Pat Hesketh, Andrew Archer, Stephen C. Bishop, Pete Kaiser, Fiona M. Tomley, David A. Hume, Adrian L. Smith, Damer P. Blake and Androniki Psifidi
- 130** *Novel Resilience Phenotypes Using Feed Intake Data From a Natural Disease Challenge Model in Wean-to-Finish Pigs*
 Austin M. Putz, John C. S. Harding, Michael K. Dyck, F. Fortin, Graham S. Plastow, Jack C. M. Dekkers and PigGen Canada
- 144** *Opportunities to Improve Resilience in Animal Breeding Programs*
 Tom V. L. Berghof, Marieke Poppe and Han A. Mulder
- 159** *Assessment of Autozygosity Derived From Runs of Homozygosity in Jinhua Pigs Disclosed by Sequencing Data*
 Zhong Xu, Hao Sun, Zhe Zhang, Qingbo Zhao, Babatunde Shittu Olasege, Qiumeng Li, Yang Yue, Peipei Ma, Xiangzhe Zhang, Qishan Wang and Yuchun Pan
- 170** *The Genetics of Life and Death: Virus-Host Interactions Underpinning Resistance to African Swine Fever, a Viral Hemorrhagic Disease*
 Christopher L. Netherton, Samuel Connell, Camilla T. O. Benfield and Linda K. Dixon
- 182** *Quantitative Trait Loci Mapping for Lameness Associated Phenotypes in Holstein–Friesian Dairy Cattle*
 Enrique Sánchez-Molano, Veysel Bay, Robert F. Smith, Georgios Oikonomou and Georgios Banos



Transcriptional Innate Immune Response of the Developing Chicken Embryo to Newcastle Disease Virus Infection

Megan A. Schilling^{1,2,3}, Robab Katani^{1,2,4}, Sahar Memari⁵, Meredith Cavanaugh⁵, Joram Buza³, Jessica Radzio-Basu¹, Fulgence N. Mpenda³, Melissa S. Deist⁶, Susan J. Lamont⁶ and Vivek Kapur^{1,2,3*}

¹ Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, United States, ² Department of Animal Science, Pennsylvania State University, University Park, PA, United States, ³ School of Life Sciences and Bio-Engineering, The Nelson Mandela African Institution of Science and Technology, Arusha, Tanzania, ⁴ Applied Biological Research Laboratory, Pennsylvania State University, University Park, PA, United States, ⁵ Department of Biology, Pennsylvania State University, University Park, PA, United States, ⁶ Department of Animal Science, Iowa State University, Ames, IA, United States

OPEN ACCESS

Edited by:

Mark S. Fife,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Jacqueline Smith,
The University of Edinburgh,
United Kingdom
Androniki Psifidi,
Royal Veterinary College,
United Kingdom

*Correspondence:

Vivek Kapur
vkapur@psu.edu

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 27 September 2017

Accepted: 09 February 2018

Published: 27 February 2018

Citation:

Schilling MA, Katani R, Memari S, Cavanaugh M, Buza J, Radzio-Basu J, Mpenda FN, Deist MS, Lamont SJ and Kapur V (2018) Transcriptional Innate Immune Response of the Developing Chicken Embryo to Newcastle Disease Virus Infection. *Front. Genet.* 9:61. doi: 10.3389/fgene.2018.00061

Traditional approaches to assess the immune response of chickens to infection are through animal trials, which are expensive, require enhanced biosecurity, compromise welfare, and are frequently influenced by confounding variables. Since the chicken embryo becomes immunocompetent prior to hatch, we here characterized the transcriptional response of selected innate immune genes to Newcastle disease virus (NDV) infection in chicken embryos at days 10, 14, and 18 of embryonic development. The results suggest that the innate immune response 72 h after challenge of 18-day chicken embryo is both consistent and robust. The expression of CCL5, Mx1, and TLR3 in lung tissues of NDV challenged chicken embryos from the outbred Kuroiler and Tanzanian local ecotype lines showed that their expression was several orders of magnitude higher in the Kuroiler than in the local ecotypes. Next, the expression patterns of three additional innate-immunity related genes, IL-8, IRF-1, and STAT1, were examined in the highly congenic Fayoumi (M5.1 and M15.2) and Leghorn (Ghs6 and Ghs13) sublines that differ only at the microchromosome bearing the major histocompatibility locus. The results show that the Ghs13 Leghorn subline had a consistently higher expression of all genes except IL-8 and expression seemed to be subline-dependent rather than breed-dependent, suggesting that the innate immune response of chicken embryos to NDV infection may be genetically controlled by the MHC-locus. Taken together, the results suggest that the chicken embryo may represent a promising model to studying the patterns and sources of variation of the avian innate immune response to infection with NDV and related pathogens.

Keywords: backyard poultry, chicken embryo, Newcastle disease virus, innate immune response, transcriptional response

INTRODUCTION

Newcastle disease virus (NDV) is one of the most important poultry pathogens worldwide, with over eighty countries in North and South America, Europe, Asia, and Africa reporting outbreaks each year (Diel et al., 2012). NDV infections manifest through a wide range of strain dependent symptoms including those within the respiratory system (coughing, sneezing, and wheezing), the nervous system (such as twisted neck, tremors, and paralysis), and the reproductive system (loss in egg production). Mortality rates may reach as high as 100% in unvaccinated flocks (Ashraf and Shah, 2014). Unsurprisingly then, NDV infections are responsible for considerable economic losses to poultry production in both developed and developing countries. For instance, the 2002–2003 NDV outbreak in California resulted in the destruction of 3 million birds and financial losses of over \$160 million (Diel et al., 2012).

Differing levels of susceptibility to NDV have been reported among commercial breeds, as well as between commercial poultry and local ecotypes; however, the mechanisms contributing to these differences remain unknown (Zhou and Lamont, 1999; Minga et al., 2004; Deist et al., 2017b). For instance, studies report that the highly inbred Fayoumi lines are less susceptible to NDV infection (and other infections) than single comb, highly inbred White Leghorn chickens (Zhou and Lamont, 1999). Similarly, backyard chickens are generally considered less susceptible to NDV and other infections than commercial chickens that are bred for high productivity. This reduced susceptibility is presumably due to their pre-sensitization with a higher frequency of natural exposure to pathogens in the scavenging environment, as well as through natural selection for hardiness to higher levels of pathogen exposure. However, considerable variation in response to NDV infection has been noted both between and within ecotypes commonly found in backyard settings (Minga et al., 2004). Another breed of chickens hypothesized to be less susceptible to disease is the Kuroiler, a dual-purpose chicken, first bred for improving both meat and egg production of backyard poultry in India. The Kuroiler, which has multi-colored feathers to help with camouflage in the wild, and thrives in backyard or scavenging environments has recently been introduced in East Africa (including Tanzania) with reports suggesting that it can coexist with, and out-produce the local chickens in Uganda and Kenya (Dessie and Getachew, 2016).

Although differences in susceptibility are observed between these breeds, the underlying immune mechanisms contributing to these differences remain unknown. With the widespread use of NDV vaccines, studies of immunity have focused on antibody production and cell-mediated responses more so than innate immunity (Ahmed et al., 2007; Kapczynski et al., 2013). One study demonstrated a rapid and robust innate response shortly after virulent NDV infection using a microarray analysis of chicken spleen tissue (Rue et al., 2011). However, the level of susceptibility to NDV was not examined in the study, hence, whether the innate immune response plays a role in enhancing immunity to NDV in poultry, and particularly in backyard poultry, is presently not understood.

Current techniques to evaluate the immune response and disease susceptibility in chickens through challenge of live birds are expensive and difficult to interpret due to confounding factors including age, natural exposure to infectious agents, the normal microflora, variability in dosing animals through natural routes of exposure, nutritional status, as well as exposure to other environmental stressors. In contrast, the chicken embryo, enclosed in the protective environment of the shell, is not only considerably cheaper, but also may be influenced by confounding environmental factors that may impact the results of challenge studies in hatched chicks.

It has been well described that the developing chicken embryo is able to produce an immune response to a pathogen prior to hatch, a feature that is widely exploited for modern large-scale poultry production with the routine administration of *in ovo* vaccination for multiple pathogens, including Marek's Disease (MD) and Infectious Bursal Disease (IBD) (Sharma and Burmester, 1984; Stone et al., 1997; Seal et al., 2000). Embryonic development occurs over 21 days and by the 10th day, the first signs of the immune system are observed. On days 11 and 12, T cells and B cells are developed, respectively, with B cell differentiation occurring after the 15th day of development. By the 18th day of embryonic development, the chicken embryo is immunocompetent and is capable of producing both an innate and adaptive response to pathogen (Davison, 2003; Ribatti, 2010; Mississippi State University Extension, 2017). Importantly, immune responses have been shown to be comparable in birds vaccinated *in ovo* or at later time points post-hatch (Sharma and Burmester, 1984; Stone et al., 1997; Gagic et al., 1999; Sharma, 1999; Steel et al., 2008). For example, vaccination with an inactivated oil emulsion NDV and Avian Influenza Virus (AIV) vaccine on the 18th day of embryonic development resulted in high seroconversion rates and antibody titers post hatch (Stone et al., 1997). An MD *in ovo* vaccine was also able to generate a four times greater level of protection than post hatch vaccination (Sharma and Burmester, 1984). Even though *in ovo* vaccination has now been applied for several decades, the mechanisms of induction of protective immunity in the chicken embryo remain poorly understood.

To begin to address this knowledge gap, we here transcriptionally profiled the innate immune response of the chicken embryo to NDV infection in both highly inbred and outbred lines. Our studies begin to demonstrate the use of the chicken embryo as a tool to examine the immune response to NDV since signatures of a consistent, breed-dependent innate immune response post NDV infection are present.

MATERIALS AND METHODS

Ethics Statement and Animal Use

Animal use protocols were approved by the Pennsylvania State University IACUC committee (protocol numbers 46395 and 47175). Specific pathogen free (SPF) eggs from White Leghorn chickens were sourced through Charles River Laboratories International, Inc. (North Franklin, CT, United States). Tanzanian local ecotype and Kuroiler hatching eggs were sourced

from Urrio Cross and Pure Breeding LTD, a local farm, in Tanzania (Tengeru, Arusha, Tanzania). Embryonated eggs from two well-defined, inbred Leghorn sublines, Ghs6 and Ghs13, as well as two inbred Fayoumi sublines, M5.1 and M15.2, from Iowa State University Poultry Farm (Ames, IA, United States) were also included in this study. Eggs were incubated (37.5°C, 55% humidity, rotating hourly) and only temporarily removed from incubation to candle for viability and perform inoculations with virus.

Virus

The lentogenic LaSota strain of NDV was kindly provided by Dr. Siba Samal at the University of Maryland, College of Veterinary Medicine (College Park, MD, United States). Titration of the virus was performed with the final titer of undiluted viral suspension of 10^7 50% egg infectious dose (EID₅₀)/mL. The viral suspension was stored at −80°C until further use.

Embryonated Chicken Egg Infection with NDV and Chicken Embryo Tissue Harvest

Embryonated eggs were inoculated at days 10, 14, or 18 of embryonic development, respectively, with 0.1 mL of the viral suspension directly deposited into the allantoic fluid. Eggs were sealed with an adhesive glue and placed back in incubators until death or removal for tissue harvest. Controls (uninfected eggs) were treated similarly with inoculation of 0.1 mL of phosphate buffered saline (PBS). Eggs were candled daily once infected and embryo (including dead embryo at that time point) were randomly selected for tissue harvest 24, 48, and 72 h post infection (hpi). The embryo selected in the group infected at 18 days and harvested 72 hpi were as close to hatch as possible without allowing the chick to hatch. Prior to harvest, the eggs were placed at 4°C for 3–4 h to avoid opening eggs with viable embryos, the harvested embryo was extracted and rinsed three times in 1X PBS before selected tissues (based on size of the embryo) were harvested (Supplementary Table S1). Tissues harvested at the 10-day infection included two sections, the body and head, as the chicken embryo was too small and underdeveloped to obtain individual tissue samples. After the 14-day infection, the spleen, heart, and liver were harvested, and after 18-day infection, the spleen, heart, liver, and lung tissues were harvested. In the experiment using the SPF White Leghorn eggs, tissues from three infected and three control embryo were harvested at each time point; for the Kuroiler and Local ecotype embryos, eight infected and eight control embryos were harvested; and for the Leghorn and Fayoumi sublines, six infected and six control were harvested in the Ghs6, Ghs13, and M5.1 and eight infected and six control embryos were harvested for M15.2. The tissues were immediately stored at −80°C prior to further use.

RNA Extraction

RNA from all tissues was extracted using the RNeasy Plus Kit (QIAGEN Inc., Germantown, MD, United States) following the recommended protocol once tissues were homogenized. Tissue homogenization was performed using a Mini-Beadbeater-96

(Biospec Products, Bartlesville, OK, United States) for 1 min. Each tube placed in the beat beater contained 100 mg of respective tissue, 600 µL of RLT Lysis buffer (provided in the RNeasy kit), and 10–15 1.5 mm silica beads (Biospec Products, Bartlesville, OK, United States). After homogenization, 600 µL of lysate was added to a clean microcentrifuge tube and centrifuged at 10,000 rpm for 3 min to remove any debris. The supernatant was transferred to the gDNA eliminator column (provided in the RNeasy kit) and the recommended protocol was followed from this point forward.

cDNA Synthesis

cDNA synthesis was performed immediately after RNA extraction using the RT² First Strand Kit (QIAGEN Inc., Germantown, MD, United States) for the samples used with the RT² Profiler Array (one control and one infected embryo at 18 days and harvested 72 h post infection) and the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Carlsbad, CA, United States) for all other samples at all other time points. Manufacturer protocols were followed using 2 µg of each respective RNA sample. cDNA was stored at −20°C.

Transcriptional Profiling of Innate and Adaptive Immune Responses

The RT² Profiler Array Chicken Innate and Adaptive Immune Responses (QIAGEN Inc., Germantown, MD, United States) was used as an initial screen of the immune response of embryo to NDV infection. The array contains a panel of 84 immune-related genes from the innate (Pattern Recognition Receptors, Cytokines, Other), adaptive (Th1, Th2, Th17, and Treg markers, T Cell Activation, Cytokines, Other), humoral, and inflammatory responses, as well as the defense response to bacteria and virus (QIAGEN, 2014). All recommended kits and procedures were followed (QIAGEN Inc., Germantown, MD, United States). A master mix was prepared using the RT² qPCR Master Mix and cDNA according to manufacturer protocols and added to each well of the 96 well plate, provided by QIAGEN Inc. qPCR was then performed using an Applied Biosystems StepOne Plus real-time PCR instrument per recommended protocols.

At the Pennsylvania State University, the PowerUp SYBR Green Master Mix (Applied Biosystems, Carlsbad, CA, United States) with the StepOne Plus System (Applied Biosystems, Carlsbad, CA, United States) was used to analyze the SPF White Leghorn, inbred Leghorn, and inbred Fayoumi samples. At the Nelson Mandela African Institution of Science and Technology, the QuantiNova SYBR Green RT-PCR Kit (QIAGEN Inc., Germantown, MD, United States) with the QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems, Carlsbad, CA, United States) was used to analyze the Kuroiler and local ecotype samples. The recommended manufacturer's protocols were followed for both kits. The differences in protocols were due to availability of reagents and machines at NM-AIST. The QuantiNova kit was recommended as most stable for shipment to Tanzania. The primers used for analysis can be found in Supplementary Table S2.

Data Analysis

Gene expression was analyzed using the $\Delta\Delta\text{Ct}$ method comparing infected ΔCt values (normalized with B-actin) with the average of negative ΔCt values (normalized with B-actin) for each gene studied. The figures were generated in R¹ using log2 fold change expression data obtained from the $\Delta\Delta\text{Ct}$ method for gene expression analysis. Statistical analysis performed between the Kuroiler and local ecotypes was performed in R using the Student's *t*-test and between the Leghorn and Fayoumi sublines using the non-parametric pairwise *t*-test (Kruskal–Wallis test) in (R Core Team, 2016).

Gene network analysis was performed using the Ingenuity Pathway Analysis (IPA) software (QIAGEN Inc., Germantown, MD, United States²) (QIAGEN, 2014). The data was uploaded to IPA and Core analysis was performed to generate the network in **Figure 2A**. From this the genes with the most gene–gene interactions and highest fold changes were selected (**Figure 2B**). The dataset was filtered for these six genes and the top diseases and biofunctions, in particular related to infectious disease and relevant to this study, were selected and the network in **Figure 2C** was generated. The same networking was performed for the four sublines (**Figure 4**).

RESULTS

Studies were performed to define the point during embryonic development when the chicken embryo is capable of producing a robust and consistent immune response to NDV, as well as to identify specific immune genes with the greatest differential expression post infection. Developing chicken embryos of various ages were infected with the LaSota strain of NDV via the allantoic fluid and tissues were harvested 24, 48, and 72 h post infection (hpi) (Supplementary Table S1).

The comparative transcriptional profile of SPF White Leghorn chicken embryos (one control and one infected at 18 days, harvested 72 hpi) was determined using the Chicken Innate and Adaptive Immune Responses RT² Profiler Array (QIAGEN Inc., Germantown, MD, United States) as an initial screen to select for immune genes in the embryo that are differentially expressed during infection since studies have not been performed previously. The upregulated genes range from 1-fold (IRF6 gene) to 754-fold (Mx1 gene) increases, and the downregulated genes range from −1-fold (IL1R1 gene) to −11-fold (CCR4 gene) decreases (Supplementary Table S3). Three of the genes with the highest increase in expression were selected for a more comprehensive analysis of the immune response in the embryo, C-C Motif Chemokine Ligand 5 (CCL5), MX Dynamin Like GTPase 1 (Mx1), and Toll-like Receptor 3 (TLR3). These three genes are also known to play a role in the innate immune response, in particular the immune response to viral infection (QIAGEN, 2014). This data is publically available at <https://doi.org/10.18113/D3H952>.

Once the three genes above were selected for further studies, the other collected embryo tissues from the SPF White Leghorns (three infected embryos and three control embryos per time point) were examined for the expression of these genes throughout the immune development of the chicken embryo (Supplementary Figure S1 and Table S1). Expression of all three immune genes was increased in the infected lung tissues at 18 days of embryonic development 72 hpi with considerably lower variation in expression than other infection time points (Supplementary Figure S1). Additionally, the Mx1 gene had a greater fold increase in the lung tissues as compared to CCL5 and TLR3 (average fold changes of 254, 4.2, and 5.6, respectively). Infection of the chicken embryo at 18 days, in particular with harvest of tissues 72 hpi, was deemed the most suitable time point for examining the immune response of the chicken embryo to NDV infection.

After determining a suitable time point to study chicken embryo immune gene expression, the analysis was broadened to include the Kuroiler and local ecotypes in Tanzania. The same three genes, CCL5, Mx1, and TLR3, were examined in the lung tissue of the Kuroiler and local ecotype embryos infected at 18 days of embryonic development and harvested 72 hpi (eight infected embryos and eight control embryos per line). Striking differences in expression of immune genes were observed between the Kuroiler and local ecotypes. The Kuroiler consistently expressed CCL5, Mx1, and TLR3 several orders of magnitude greater (164-, 19,816-, and 21.8-fold increase, respectively) than the local ecotype embryos (4.2-, 4.7-, and 1.8-fold increase, respectively) (**Figure 1**; data for individual replicates³).

Pathway analysis using the Ingenuity Pathway Analysis (IPA) software (QIAGEN Inc., Germantown, MD, United States²) was performed using the RT² Profiler Array data (expression fold change cutoff = 2) to determine important pathways and networks in the response to NDV in the chicken embryo (QIAGEN, 2014). The top canonical pathways included the role of pattern recognition receptors in recognition of bacteria and virus, iNOS Signaling, communication between innate and adaptive immune cells, and Toll-like receptor signaling (each pathway – *p*-values < 0.0001). The top regulator effector network was the antiviral response. The identified associated network diseases and functions including infectious disease, the cell-mediated immune response, and organismal injury, were selected based on relevance to the immune response and merged to generate a network map of the gene–gene interactions from the RT² Profiler Array, removing genes that were not included in the array (**Figure 2A**). From this, an additional three genes, Interleukin-8 (IL-8), Interferon Regulatory Factor 1 (IRF1) and Signal Transducer and Activator of Transcription 1 (STAT1), were selected for future analysis since they were all highly increased in the infected chicken embryo as well as being interconnected in immune response networks (**Figure 2B**). Core analysis was performed with filtering for these genes to determine important networks of relevance to this study with NDV (**Figure 2C**). Remarkably,

¹<https://www.r-project.org/>

²<https://targetexplorer.ingenuity.com/>

³<https://doi.org/10.18113/D3H952>

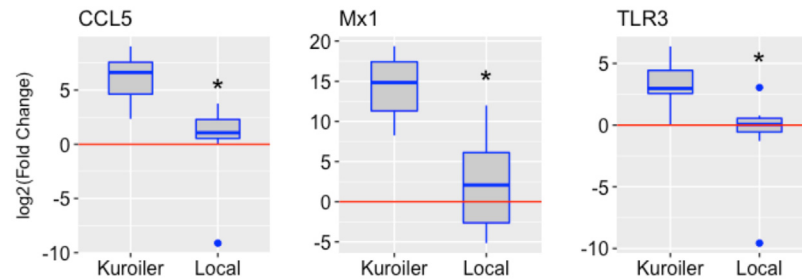


FIGURE 1 | Expression of CCL5, Mx1, and TLR3 in NDV-infected embryos from the Kuroiler and local ecotypes in Tanzania. Differences in the log2 of the fold change are shown in the boxplots. The red line signifies the cutoff between upregulation and downregulation of the genes. Gene expression is significantly different between the Kuroiler and Local ecotypes ($n = 8$ infected, 8 control per line) for all of the genes studied (* p -values < 0.001).

the diseases and disorders associated with the expression of these six genes include NDV infection, RNA virus infection and replication, Paramyxovirus replication, lung infection, Measles virus infection and replication (a paramyxovirus), and Influenza A virus infection (a RNA virus of the respiratory tract), and susceptibility (p -values < 0.0001 for all diseases and disorders listed) (QIAGEN, 2014). However, we note that even though IPA analysis has been previously applied to understand patterns of gene expression in chickens, the pathways have been validated primarily using human/mouse/rat data, and it is also recommended to use at least 200 genes for pathways analyses, and hence, the results should be independently replicated (Krämer et al., 2014).

Next, the immune response profiles of 18-day NDV-infected embryos of the well-defined, highly congenic Fayoumi (M5.1 and M15.2) and Leghorn (Ghs6 and Ghs13) sublines were characterized (six infected embryos and six control embryos per subline – Ghs6, Ghs13, M5.1; eight infected and six control for the M15.2 subline). The congenic Fayoumi and Leghorn lines are highly inbred commercial poultry with inbreeding coefficients of 0.99, differing within each line only at the microchromosome bearing the MHC (Zhou and Lamont, 1999; Deist et al., 2017b). The Fayoumi was described as less susceptible to infectious disease than the Leghorn (Zhou and Lamont, 1999; Cheeseman, 2007). One study showed the Fayoumi line (M15.2) to be less susceptible to NDV than the Leghorn line (Ghs6) based on viral load post-challenge in ocular secretions at 6 (but not 2) days post infection (Hermann et al., 2016; Deist et al., 2017b), and have also identified differentially expressed genes in the trachea and lung transcriptomes of these two highly inbred chicken lines (Deist et al., 2017a,b). The results of our current investigation show striking differential expression of select genes in a subline-dependent manner with Ghs13 having consistently higher expression of all genes except IL-8 (Figure 3; data for individual replicates⁴). Importantly, the results show that the gene expression differences seem to be subline-dependent even more so than breed dependent.

Similar analysis was performed, as in Figure 2C, with the average fold expression differences between challenged

and control lung tissue in the inbred Leghorn and inbred Fayoumi sublines using IPA (QIAGEN Inc., Germantown, MD, United States⁵). Interestingly, the high responders, Ghs13 and M5.2 (Figures 4A,B, respectively), mimic similar patterns as the RT² Profiler Array analysis (Figure 3). The low responders, Ghs6 and M15.2 (Figures 4C,D, respectively), differ especially due to the downregulation of one or more of these genes, STAT1, CCL5, and IRF1, was predicted to increase viral replication and infection in the Ghs6 and M15.2 sublines (QIAGEN, 2014).

DISCUSSION

Until now, most studies performed to examine the immune response of chickens and other avian species have been through expensive and laborious *in vivo* experiments or through *in vitro* cell culture experiments lacking host–pathogen interactions. The animal studies, requiring enhanced biosecurity facilities, are influenced by multiple confounding factors producing results that are sometimes difficult to interpret. Therefore, despite current work to examine the chicken immune response to NDV, a knowledge gap remains in the mechanisms of the immune response and the effects they have on susceptibility of chickens to infection. In the present study, we utilized the developing chicken embryo as a controlled and inexpensive approach to evaluate the innate immune mechanisms in response to NDV infection in chickens. Through these studies we have found evidence that the chicken embryo is capable of producing a robust signature of the immune response, in particular the innate immune response, to NDV infection, at least after day 10 of embryonic development. In particular, the suitability of the chicken embryo for immune investigation was confirmed through gene expression studies where variation in gene expression was greatly reduced at 18-day infection as compared to earlier infection time points (Supplementary Figure S1). Since the chicken embryo is immunocompetent and the respiratory system almost fully developed a few days prior to hatch, examining the response of the chicken embryo to respiratory viruses such as NDV may be possible without the introduction of other confounding variables post-hatch.

⁴<https://doi.org/10.18113/D3H952>

⁵<https://targetexplorer.ingenuity.com/>

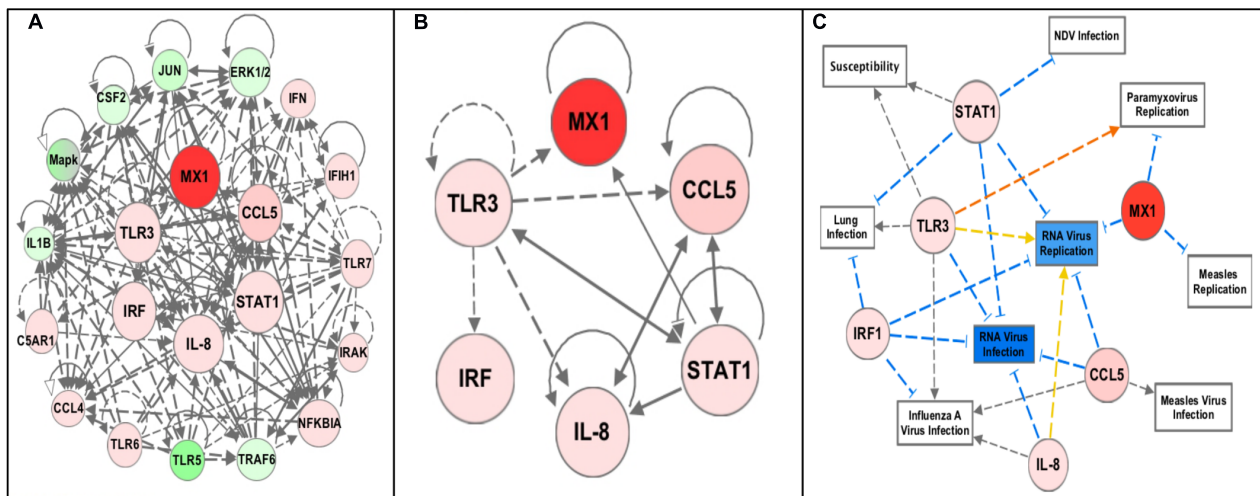


FIGURE 2 | Pathway analysis of the RT² Profiler Array gene expression data using Ingenuity Pathway Analysis (IPA) software. The red colored circles represent relative increases in expression of that gene in the infected chicken embryo lung tissues versus control and green circles represent downregulation. Dashed lines indicate direct relationships and solid lines indicate indirect relationships. The blue lines represent inhibition, orange represents activation, and yellow represents inconsistent findings with downstream molecules. **(A)** Networks associated with diseases and functions of the differentially expressed genes from the profiler array (Infectious Diseases, Cell-mediated Immune Responses, Lymphoid Tissue Development, and Organismal Injury) were mapped together. **(B)** The six highly upregulated and interconnected genes (MX1, CCL5, TLR3, STAT1, IL-8, and IRF1) from the array were selected for future analysis. **(C)** Filtering of the data to include only these six genes revealed diseases and functions relevant to NDV infection, including NDV infection, Paramyxovirus replication, RNA virus infection and replication, susceptibility to infection, lung infection, Measles virus infection and replication, and Influenza A virus infection (all p -values < 0.0001).

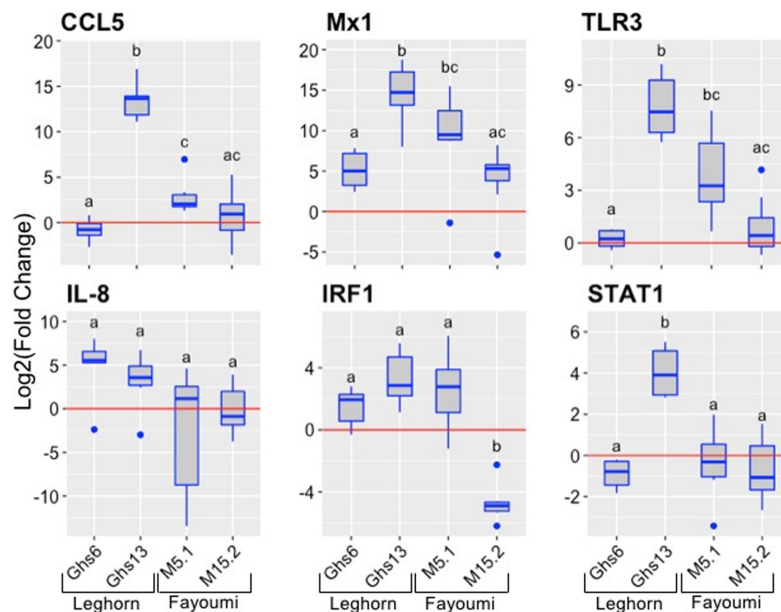


FIGURE 3 | Expression of immune genes in the Leghorn (Ghs6 and Ghs13) and Fayoumi (M5.1 and M15.2) sublines. The six immune genes were examined in the lung tissues of each subline (Ghs6, Ghs13, M5.1 – $n = 6$ infected and 6 controls per subline; M15.2 – $n = 8$ infected and 6 control). The red line signifies the cutoff between upregulation and downregulation of the genes. Statistical analysis was performed according to the materials and methods data analysis and significant p -values noted by the superscripts (p -value < 0.05) can be found in Supplementary Table S4.

Using this more controlled approach to examine the chicken immune response, we were able to demonstrate that the innate immune response to NDV in the chicken embryo appeared to be breed- (Figure 1) and/or subline-dependent (Figure 3), with the

possibility of having a relation to the MHC type demonstrated by differences in the inbred Fayoumi and inbred Leghorn sublines, which only differ in the MHC-bearing chromosome. If the gene expression is, in fact, MHC type dependent, this

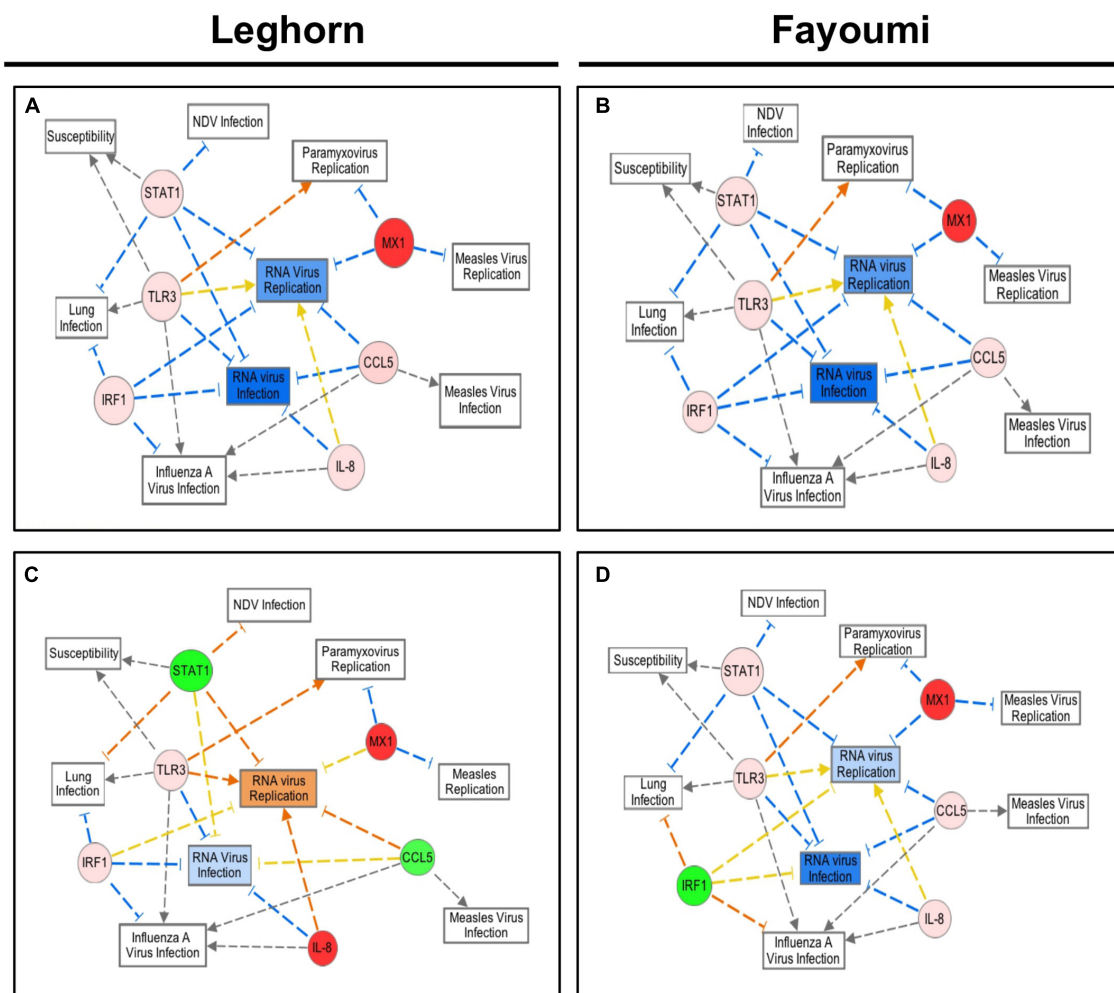


FIGURE 4 | Network Analysis of the six selected immune genes in the Leghorn and Fayoumi sublines using IPA software. The data used for this analysis is from the gene expression data from each Fayoumi and Leghorn subline in **Figure 3**. The red colored circles represent relative increases in expression of that gene in the infected chicken embryo lung tissues versus control and green circles represent downregulation. Dashed lines indicate direct relationships and solid lines indicate indirect relationships. The blue lines represent inhibition, orange represents activation, and yellow represents inconsistent findings with downstream molecules. Ghs13 (**A**) and M5.1 (**B**) the highest responders have similar patterns in the network analysis, and Ghs6 (**C**) and M15.2 (**D**) show deviations from this network due to the downregulation of some genes.

could explain some of the large within breed variation in expression demonstrated by the Tanzanian local ecotypes and the Kuroiler. These differences in the transcriptional response to NDV infection also demonstrate a possible relation between the innate immune gene expression and level of susceptibility of a particular line, since previous reports demonstrate the Ghs6 Leghorn subline and M15.2 Fayoumi subline differ in the level of susceptibility to NDV (Deist et al., 2017b). This study also showed IL-8 and Mx1 to be differentially expressed between the two sublines studied (Deist et al., 2017b). A more recent investigation of differential gene expression in the lung transcriptome of these same chicken lines in response to NDV infection also, demonstrated an activation of the IL-8 pathway in the resistant Fayoumi line, M15.2 (Deist et al., 2017a). While these previous studies used RNA-Seq based approaches and longer time-points (days rather than hours) post-infection and performed these

studies in hatched chicks and are hence not directly comparable, the similar patterns of expression of at least some of the loci observed in the chick embryo model are encouraging, and need to be confirmed in future investigations.

Multiple other studies have each discovered at least one of the six genes (Mx1, CCL5, TLR3, IL-8, IRF1, and STAT1) studied here with elevated expression levels in the chicken particularly in response to NDV, AIV, and IBV demonstrating the key role of these innate immune genes in the chicken's response to pathogen (Heidari et al., 2010; Rue et al., 2011; Matulova et al., 2013; Cheng et al., 2014; Kang et al., 2016; Ranaware et al., 2016; Deist et al., 2017b). Although the network analysis was only performed with small numbers of genes, relevant pathways to our study were generated. Most surprising were the pathways, top diseases, and disorders associated with the genes including NDV infection, paramyxovirus replication, and lung infection since

they are closely associated with the study of NDV infection in the chicken embryo lung directly. Another significant function of relevance to this study was susceptibility to infection. Since determining the feasibility of using the chicken embryo as a new approach to explore the immune response and susceptibility to NDV infection is an ultimate goal of our project, revealing a significant involvement of STAT1 and TLR3 in this response is promising for future studies. Other diseases and disorders include RNA virus infection and replication (NDV is an RNA virus), Measles virus infection and replication (a paramyxovirus, like NDV), and Influenza A virus infection (an RNA virus of the respiratory tract) (QIAGEN, 2014). In most cases, TLR3, STAT1, and IRF1 are involved in signaling pathways that lead to stimulation of the end targets, CCL5, Mx1, and IL-8, which then go on to stimulate other immune related cells or pathways (QIAGEN, 2014). Multiple studies have recognized the Type-I Interferon (IFN) inducible response, protecting cells against invading viral pathogens, as important in the innate immune response in chickens, and the six genes examined in this study are involved in that response (Xing et al., 2008; Schoggins et al., 2011; Ranaware et al., 2016). The six genes studied here have been shown to be important in the chicken immune response to viral pathogens, whether it is NDV, AIV, or IBV (Ko et al., 2002; Benfield et al., 2008; Sartika et al., 2011; Cong et al., 2013; Barjesteh et al., 2014; Cheng et al., 2014; Dou et al., 2014; Fulton et al., 2014; Ranaware et al., 2016; Deist et al., 2017a,b), which demonstrates how the chicken embryo immune response starts to mimic that of hatched chickens, validating the use of the chicken embryo for future studies of the immune response to avian pathogens.

The innate immune response is a complex and interconnected network that is dependent on many factors. Although the six genes examined in this study are involved in the innate immune response and there are direct interactions between these genes, further validation of the response, including a larger set of innate immune genes, is needed in order to obtain a better understanding of the innate immune mechanisms influencing the chicken embryo and chicken's response to NDV infection. It is interesting to note the differences seen in the pathway analysis from the Leghorn and Fayoumi sublines and a possible role for these genes in determining the level of susceptibility of a particular line to NDV infection. Most interestingly, the Ghs6 subline, previously found to be susceptible to NDV, differs from the other sublines pathway analysis by activating RNA virus replication rather than inhibiting it (QIAGEN, 2014; Deist et al., 2017b). Since only the Ghs6 and M15.2 sublines have been previously characterized for the level of susceptibility to NDV, expanding research in this area to include the other two sublines may provide a better insight into the use of the chicken embryo immune response as a tool to screen for the level of susceptibility to NDV of a particular line, and the role of the MHC complex. It is especially intriguing since the high responders, Ghs6 and M5.1, differ in the network analysis from the low responder, Ghs13. A more comprehensive study of the innate immune response of the different chicken lines, including a larger set of genes and larger sample sizes,

and correlating these responses to the innate immune response and level of susceptibility of hatched chicks, using phenotypic characteristics such as viral load, mean death time, viral shedding, and antibody titers post NDV infection, will help provide insight into innate immune mechanisms of susceptibility to NDV. Other important future studies include assessing whether similar patterns in response using more pathogenic field strains which is especially important in regards to the response in backyard poultry.

Newcastle disease virus can be particularly devastating for farmers, especially smallholder farmers in Sub-Saharan Africa, due to high mortality rates in unvaccinated flocks. Uncovering innate immune mechanisms related to the level of susceptibility to NDV would have major impacts on productivity for these farmers by informing breeding strategies to produce more robust chickens. We note that the potential use of the chicken embryo as a model provides a framework for future studies of the development of the innate immune response of chickens to NDV (and other pathogens) allowing for screening of large numbers of birds to uncover genetic markers for both disease resistance/susceptibility and productivity traits, such as egg or meat production.

AUTHOR CONTRIBUTIONS

MS, VK, JB, and FM designed the study. MS, SM, and MC performed the laboratory experiments. MS, RK, SM, and MC performed the data analysis. MS wrote the manuscript. RK, JR-B, MD, and SL critically revised the content. All authors read and approved the final manuscript.

FUNDING

This research was supported by grant (OPP1083453) from the Bill and Melinda Gates Foundation (to VK and MS) and USAID Feed the Future Innovation Lab for Genomics to Improve Poultry and Hatch project #5357, USDA NIFA 2013-38420-20496 (to SL).

ACKNOWLEDGMENTS

The authors thank the Kapur Lab, especially Dr. Lingling Li, for her guidance and support throughout the study. They thank Dr. Huaguang Lu and Dr. Suresh Kuchipudi and their lab groups for their guidance and technical assistance with the chicken embryo experiments. They also thank Dr. Beatus Lyimo for his support during experimentation at the Nelson Mandela African Institution of Science and Technology in Arusha, TZ.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00061/full#supplementary-material>

REFERENCES

- Ahmed, K. A., Saxena, V. K., Ara, A., Singh, K. B., Sundaresan, N. R., Saxena, M., et al. (2007). Immune response to Newcastle disease virus in chicken lines divergently selected for cutaneous hypersensitivity. *Int. J. Immunogenet.* 34, 445–455. doi: 10.1111/j.1744-313X.2007.00722.x
- Ashraf, A., and Shah, M. (2014). Newcastle disease: present status and future challenges for developing countries. *Afr. J. Microbiol. Res.* 5, 411–416. doi: 10.5897/AJMR2013.6540
- Barjesteh, N., Behboudi, S., Brisbin, J. T., Villanueva, A. I., Nagy, E., and Sharif, S. (2014). TLR ligands induce antiviral response in chicken macrophages. *PLoS One* 9:e105713. doi: 10.1371/journal.pone.0105713
- Benfield, C. T. O., Lyall, J. W., Kochs, G., and Tiley, L. S. (2008). Asparagine 631 variants of the chicken Mx protein do not inhibit influenza virus replication in primary chicken embryo fibroblasts or in vitro surrogate assays. *J. Virol.* 82, 7533–7539. doi: 10.1128/JVI.00185-08
- Cheeseman, J. H. (2007). *Avian Immunology, Immunogenetics, and Host Immune Response to Salmonella enterica Serovar Enteritidis Infection in Chickens*. Doctoral dissertation, Iowa State University, Ames, IA.
- Cheng, J., Sun, Y., Zhang, X., Zhang, F., Zhang, S., Yu, S., et al. (2014). Toll-like receptor 3 inhibits Newcastle disease virus replication through activation of pro-inflammatory cytokines and the type-1 interferon pathway. *Arch. Virol.* 159, 2937–2948. doi: 10.1007/s00705-014-2148-6
- Cong, F., Liu, X., Han, Z., Shao, Y., Kong, X., and Liu, S. (2013). Transcriptome analysis of chicken kidney tissues following coronavirus avian infectious bronchitis virus infection. *BMC Genomics* 14:743. doi: 10.1186/1471-2164-14-743
- Davison, T. F. (2003). The immunologists' debt to the chicken. *Br. Poult. Sci.* 44, 6–21. doi: 10.1080/0007166031000085364
- Deist, M. S., Gallardo, R. A., Bunn, D. A., Dekkers, J. C., Zhou, H., and Lamont, S. J. (2017a). Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. *BMC Genomics* 18:989. doi: 10.1186/s12864-017-4380-4
- Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., Dekkers, J. C. M., Zhou, H., et al. (2017b). Novel mechanisms revealed in the trachea transcriptome of resistant and susceptible chicken lines following infection with Newcastle disease virus. *Clin. Vaccine Immunol.* 24:e00027-17. doi: 10.1128/CI.00027-17
- Dessie, T., and Getachew, F. (2016). *The Kuroiler Breed. African Chicken Genetic Gains Fact Sheet 2*. Nairobi: ILRI.
- Diel, D. G., Susta, L., Garcia, S. C., Killian, M. L., Brown, C. C., Miller, P. J., et al. (2012). Complete genome and clinicopathological characterization of a virulent Newcastle disease virus isolate from South America. *J. Clin. Microbiol.* 50, 378–387. doi: 10.1128/JCM.06018-11
- Dou, L., Liang, H.-F., Geller, D. A., Chen, Y.-F., and Chen, X.-P. (2014). The regulation role of interferon regulatory factor-1 gene and clinical relevance. *Hum. Immunol.* 75, 1110–1114. doi: 10.1016/j.humimm.2014.09.015
- Fulton, J. E., Arango, J., Ali, R. A., Bohorquez, E. B., Lund, A. R., Ashwell, C. M., et al. (2014). Genetic variation within the Mx gene of commercially selected chicken lines reveals multiple haplotypes, recombination, and a protein under selective pressure. *PLoS One* 9:e108054. doi: 10.1371/journal.pone.0108045
- Gagic, M., St Hill, C. A., and Sharma, J. M. (1999). In ovo vaccination of specific-pathogen-free chickens with vaccines containing multiple agents. *Avian Dis.* 43, 293–301. doi: 10.2307/1592620
- Heidari, M., Sarson, A. J., Huebner, M., Sharif, S., Kireev, D., and Zhou, H. (2010). Marek's disease virus-induced immunosuppression: array analysis of chicken immune response gene expression profiling. *Viral Immunol.* 23, 309–319. doi: 10.1089/vim.2009.0079
- Hermann, M. S., Gallardo, R., Bunn, D. A., Zhou, H., and Lamont, S. J. (2016). *Do Two Distinct Chicken Lines Differ in Their Response to Newcastle Disease Virus?* Animal Industry Report AS 662, ASL R3098. Ames, IA: Iowa State University.
- Kang, Y., Feng, M., Zhao, X., Dai, X., Xiang, B., Gao, P., et al. (2016). Newcastle disease virus infection in chicken embryonic fibroblasts but not duck embryonic fibroblasts is associated with elevated host innate immune response. *Virol. J.* 13:41. doi: 10.1186/s12985-016-0499-1
- Kapczynski, D. R., Alfonso, C. L., and Miller, P. J. (2013). Immune responses of poultry to Newcastle disease virus. *Dev. Comp. Immunol.* 41, 447–453. doi: 10.1016/j.dci.2013.04.012
- Ko, J.-H., Jin, H.-K., Asano, A., Tkada, A., Ninomiya, A., Kida, H., et al. (2002). Polymorphisms and the differential antiviral activity of chicken Mx gene. *Genome Res.* 12, 595–601. doi: 10.1101/gr.210702
- Krämer, A., Green, J., Pollard, J. Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530.
- Matulova, M., Varmuzova, K., Sisak, F., Havlickova, H., Babak, V., Stejskal, K., et al. (2013). Chicken innate immune response to oral infection with *Salmonella enterica* serovar Enteritidis. *Vet. Res.* 44:37. doi: 10.1186/1297-9716-44-37
- Minga, U. M., Msoffe, P. L., and Gwakisa, P. S. (2004). "Biodiversity (variation) in disease resistance in pathogens within rural chicken populations," in *Proceedings of the International Health Network for Family Poultry (INFD). World Poultry Congress*, Istanbul.
- Mississippi State University Extension (2017). *Stages in Chick Embryo Development*. Available at: <http://extension.msstate.edu/content/stages-chick-embryo-development> [accessed March 28, 2017].
- QIAGEN (2014). *Innate and Adaptive Immune Responses PCR Array*. Available at: http://www.sabiosciences.com/rt_pcr_product/HTML/PAGG-052Z.html
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ranaware, P. B., Mishra, A., Vijayakumar, P., Gandhale, P. N., Kumar, H., Kulkarni, D. D., et al. (2016). Genome wide host gene expression analysis in chicken lungs infected with avian influenza viruses. *PLoS One* 11:e0153671. doi: 10.1317/journal.pone.0153671
- Ribatti, D. (2010). *The Chick Embryo Chorioallantoic Membrane in the Study of Angiogenesis*. Berlin: Springer Science & Business Media. doi: 10.1007/978-90-481-3845-6
- Rue, C. A., Susta, L., Cornax, I., Brown, C. C., Kapczynski, D. R., Suarez, D. L., et al. (2011). Virulent Newcastle disease virus elicits a strong innate immune response in chickens. *J. Gen. Virol.* 92, 931–939. doi: 10.1099/vir.0.025486-0
- Sartika, T., Sulandari, S., and Zein, M. S. A. (2011). Selection of Mx gene genotype as genetic marker for avian influenza resistance in Indonesian native chickens. *BMC Proc.* 4(Suppl. 4):S37. doi: 10.1186/1753-6561-5-S4-S37
- Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., et al. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472, 481–485. doi: 10.1038/nature09907
- Seal, B. S., King, D. J., and Sellers, H. S. (2000). The avian response to Newcastle disease virus. *Dev. Comp. Immunol.* 24, 257–268. doi: 10.1016/S0145-305X(99)00077-4
- Sharma, J. M. (1999). Introduction to poultry vaccines and immunity. *Adv. Vet. Med.* 41, 481–494. doi: 10.1016/S0065-3519(99)80036-6
- Sharma, J. M., and Burmester, B. R. (1984). Disease control in avian species by embryonal vaccination. USA Patent Application 391065.
- Steel, J., Burmakina, S. V., Thomas, C., Spackman, E., Garcia-Sastre, A., Swayne, D. E., et al. (2008). A combination in-ovo vaccine for avian influenza virus and Newcastle disease virus. *Vaccine* 26, 522–531. doi: 10.1016/j.vaccine.2007.11.032
- Stone, H., Mitchell, B., and Brugh, M. (1997). In ovo vaccination of chicken embryos with experimental Newcastle disease and avian influenza oil-emulsion vaccines. *Avian Dis.* 41, 856–863. doi: 10.2307/1592339
- Xing, Z., Cardona, C. J., Li, J., Dao, N., Tran, T., and Andrada, J. (2008). Modulation of the immune responses in chickens by low-pathogenicity avian influenza virus H9N2. *J. Gen. Virol.* 89(Pt 5), 1288–1299. doi: 10.1099/vir.0.83362-0
- Zhou, H., and Lamont, S. J. (1999). Genetic Characterization of biodiversity in highly inbred chicken lines by microsatellite markers. *Anim. Genet.* 30, 256–264. doi: 10.1046/j.1365-2052.1999.00505.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Schilling, Katani, Memari, Cavanaugh, Buza, Radzio-Basu, Mpenda, Deist, Lamont and Kapur. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Gene Expression Response to Sea Lice in Atlantic Salmon Skin: RNA Sequencing Comparison Between Resistant and Susceptible Animals

Diego Robledo^{1*}, Alejandro P. Gutiérrez¹, Agustín Barria², José M. Yáñez^{2,3†} and Ross D. Houston^{1*†}

¹ The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, United Kingdom, ² Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile, ³ Aquainnovo S.A., Puerto Montt, Chile

OPEN ACCESS

Edited by:

Mark S. Fife,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Filippo Biscarini,
Consiglio Nazionale delle Ricerche
(CNR), Italy
Dirk-Jan De Koning,
Swedish University of Agricultural
Sciences, Sweden

*Correspondence:

Diego Robledo
Diego.Robledo@roslin.ed.ac.uk
Ross D. Houston
ross.houston@roslin.ed.ac.uk

[†] These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 14 May 2018

Accepted: 11 July 2018

Published: 03 August 2018

Citation:

Robledo D, Gutiérrez AP, Barria A,
Yáñez JM and Houston RD (2018)
Gene Expression Response to Sea
Lice in Atlantic Salmon Skin: RNA
Sequencing Comparison Between
Resistant and Susceptible Animals.
Front. Genet. 9:287.
doi: 10.3389/fgene.2018.00287

Sea lice are parasitic copepods that cause large economic losses to salmon aquaculture worldwide. Frequent chemotherapeutic treatments are typically required to control this parasite, and alternative measures such as breeding for improved host resistance are desirable. Insight into the host–parasite interaction and mechanisms of host resistance can lead to improvements in selective breeding, and potentially novel treatment targets. In this study, RNA sequencing was used to study the skin transcriptome of Atlantic salmon (*Salmo salar*) parasitized with sea lice (*Caligus rogercresseyi*). The overall aims were to compare the transcriptomic profile of skin at louse attachment sites and “healthy” skin, and to assess differences in gene expression response between animals with varying levels of resistance to the parasite. Atlantic salmon pre-smolts were challenged with *C. rogercresseyi*, growth and lice count measurements were taken for each fish. 21 animals were selected and RNA-Seq was performed on skin from a louse attachment site, and skin distal to attachment sites for each animal. These animals were classified into family-balanced groups according to the traits of resistance (high vs. low lice count), and growth during infestation. Overall comparison of skin from louse attachment sites vs. healthy skin showed that 4,355 genes were differentially expressed, indicating local up-regulation of several immune pathways and activation of tissue repair mechanisms. Comparison between resistant and susceptible animals highlighted expression differences in several immune response and pattern recognition genes, and also myogenic and iron availability factors. Components of the pathways involved in differential response to sea lice may be targets for studies aimed at improved or novel treatment strategies, or to prioritize candidate functional polymorphisms to enhance genomic selection for host resistance in commercial salmon breeding programs.

Keywords: *Caligus rogercresseyi*, *Salmo salar*, aquaculture, disease, parasite, RNA-Seq, host–parasite, differential expression

INTRODUCTION

Aquaculture is currently the fastest growing food industry (Food and Agriculture Organization of the United Nations, 2016) and is essential to meet increasing global demands for fish. However, the sustainability and prolonged success of any farming industry depends on effective disease prevention and control, and this tends to be particularly challenging for aquaculture. The aquatic

environment and high stock density can expedite pathogen spread, which has historically resulted in periodic mass mortality events (Lafferty et al., 2015; Food and Agriculture Organization of the United Nations, 2017) and ongoing challenges in disease prevention and control. While biosecurity measures, vaccination, nutrition, and medicines all play vital roles for several diseases, selective breeding to produce more resistant and tolerant aquaculture stocks is rapidly becoming a key component of the battle to prevent these outbreaks (Yáñez et al., 2014a; Palaiokostas et al., 2016).

Sea lice, ectoparasites of the family Caligidae, are one of the major disease problems that the aquaculture industry is facing, and specifically for salmon farming. Atlantic salmon (*Salmo salar*) is the most important species in aquaculture with a production value of 14.7 billion US dollars in 2014 (Food and Agriculture Organization of the United Nations, 2016), therefore control of sea lice is a primary goal for the industry. Sea lice-related economic losses to worldwide salmonid aquaculture were estimated at ~430 million USD per annum (Costello, 2009). Two lice species present the primary concerns for salmon farming: primarily *Lepeophtheirus salmonis* in the Northern Hemisphere and *Caligus rogercresseyi* in the Southern Hemisphere (Johnson et al., 2004). These copepods parasitize salmon during the marine phase of the lifecycle by attaching to their skin or fins, and feeding on the blood and tissue. This leads to open wounds which can facilitate the entry of other pathogens. The impaired growth and secondary infections cause significant negative animal welfare and economic impact (Frazer et al., 2012). Despite extensive use of both chemical (i.e., hydrogen peroxide, emamectin benzoate, organophosphates, pyrethroids, or benzoyl ureas) and non-chemical treatments (i.e., fresh water bath) to control sea lice, their negative impact on salmon aquaculture has increased in the past years (Torrissen et al., 2013), and various sea lice populations have been reported to be resistant to the most common chemicals available for therapeutic control, such as emamectin benzoate, azamethiphos (organophosphate), deltamethrin (pyrethroid), and even hydrogen peroxide (Aaen et al., 2015). Therefore, alternative methods to control sea lice are currently being studied, including the use of probiotics to reduce salmon attractiveness for sea lice (Jodaa et al., 2016) or cohabitation with lice-eating species (Imsland et al., 2014; Leclercq et al., 2014).

Knowledge of the interaction between salmon and sea lice can help devise more effective prevention and treatment strategies. Therefore, a lot of effort has been put in characterizing the host response to sea lice infestation (reviewed in Fast, 2014). Interestingly, the outcome of infestation varies for different salmonid species (Johnson and Albright, 1992), with coho salmon (*Oncorhynchus kisutch*) showing rapid inflammatory response and epithelial hyperplasia, leading to parasite encapsulation and more than 90% reduction in lice loads (Fast, 2014). In comparison, Atlantic salmon (*S. salar*) is highly susceptible to sea lice infestation and seemingly cannot mount a fully effective immune response (Fast, 2014). Comparative transcriptomics has shown that iron sequestration, increased expression of pattern recognition receptors such as c-type lectins and up-regulation of pro-inflammatory cytokines such as interleukin-1 β

are observed in salmon species resistant to sea lice (Sutherland et al., 2014). Interleukin-1 β has also been implicated in successful responses to sea lice in other salmonid species such as pink salmon (*Oncorhynchus gorbuscha*) and coho salmon (Braden et al., 2012, 2015; Sutherland et al., 2015), and recently immunostimulant feeds up-regulating interleukin-1 β in skin and spleen have shown some promising results to boost Atlantic salmon resistance to sea lice (Sutherland et al., 2017). While these studies have mainly focused on *L. salmonis*, similar findings have been observed in *C. rogercresseyi* infestation. For instance, comparative analyses of Atlantic and coho salmon parasitized with *C. rogercresseyi* showed that despite both showing up-regulation of pro-inflammatory genes, the response was highly specific, characterized in coho by an activation of the TH1 response (Valenzuela-Muñoz et al., 2016). Another study linked iron sequestration and depletion mechanisms to the Atlantic salmon immune response to *C. rogercresseyi* (Valenzuela-Muñoz et al., 2017).

A promising and potentially complementary approach to existing control measures is to exploit natural genetic variation in farmed salmon populations to breed stocks with enhanced resistance to the parasite. The presence of significant genetic variation for resistance to *C. rogercresseyi*, with heritability values ranging between 0.1 and 0.34, demonstrates the feasibility of improving this trait by selective breeding in Atlantic salmon (Lhorente et al., 2014; Yáñez et al., 2014b). Current evidence indicates that host resistance to sea lice in Atlantic salmon has a highly polygenic genetic basis, with little evidence for major QTL (Ødegård et al., 2014; Gharbi et al., 2015; Correa et al., 2016, 2017; Tsai et al., 2016). Therefore, genomic selection using genome-wide markers to predict lice resistant breeding values has been widely applied in commercial Atlantic salmon breeding programs, with a relative advantage compared to pedigree selection of 10–27% (Tsai et al., 2016; Correa et al., 2017). Understanding the underlying functional basis of genetic resistance to sea lice can lead to improved methods of selective breeding. For example, incorporating functional variants into genomic prediction models could help improve prediction accuracy, in particular for cross-population prediction (MacLeod et al., 2016). Functional annotation of reference genomes is pertinent to this process, and the emerging Functional Annotation of All Salmonid Genomes (FAASG) project (Macqueen et al., 2017) is aiming to improve genome annotation for Atlantic salmon (among other species). Further, the discovery of putatively causative genes and variants could, in the near future, lead to their introduction into populations or species where it has never been present through the use of genome editing, for example using CRISPR-Cas9 technology, which has been successfully applied in salmon to knockout two genes related to pigmentation (tyrosinase and solute carrier family 45 member 2) and the *dnd* (dead end) gene, producing albino (Edvardsen et al., 2014), and germ cell-free salmon (Wargelius et al., 2016), respectively.

Expression differences between Atlantic salmon resistant and susceptible families in response to *L. salmonis* for 32 immune genes suggested that resistant fish are better at avoiding immunosuppression (Holm et al., 2015). The same study found

suggestive evidence that physical tissue barrier such as enhanced mucus production does not contribute to resistance (Holm et al., 2015). However, to our knowledge, the functional basis of genomic resistance to sea lice in Atlantic salmon has not been studied on a genome-wide scale, nor it has been explored in response to *C. rogercresseyi*. RNA sequencing can provide a first layer toward a holistic view of the host response to parasite infection, which in turn can highlight specific genes, pathways, and networks involved in the host–parasite interaction. RNA sequencing can also be used to identify single nucleotide polymorphisms in transcribed regions, and to assess the putative impact of those genetic markers on transcript and protein function. The effect of these markers on gene expression (and ultimately host resistance) can be assessed by allelic-specific expression or expression QTL studies, leading to a shortlist of candidate functional variants.

The overall aims of the current study were to compare the transcriptome profile of salmon skin at louse attachment sites and “healthy” skin (from the same fish), and to evaluate differences in these profiles between animals with varying levels of resistance to the parasite. To achieve this, challenged animals were classified into family-balanced groups according to resistance (based on high vs. low lice count) and growth during infestation, and RNA sequencing was performed on individual samples. By comparing resistant vs. susceptible samples, genes and pathways related to local immune response and host resistance were identified, and their potential role discussed.

MATERIALS AND METHODS

Experimental Design

2,668 Atlantic salmon (*S. salar*) pre-smolts (average weight: 136 g) from 104 families from the breeding population of AquaInnovo (Salmones Chaicas, Xth Region, Chile), were experimentally challenged with *C. rogercresseyi* (chalimus II–III). This population will be used for a future study on sea lice resistance genetic architecture and genomic selection. Briefly, infestation with the parasite was carried out by using 13–24 copepodids per fish and stopping the water flow for 6 h after infestation. Eight days after the infestation fish were euthanized and fins from each fish were collected and fixed for processing and lice counting. 42 samples from 21 fish from 6 different families (2–5 fish per family) were selected for RNA sequencing (Supplementary File S1) based on the traits of interest (number of sea lice attached to their fins and growth during challenge). Skin samples (both from attachment sites and health skin) were obtained from each animal and stored in RNeasy lysis buffer at 4°C for 24 h, and then at –20°C until RNA extraction for sequencing.

RNA Extraction and Sequencing

For all the 42 samples a standard TRI Reagent RNA extraction protocol was followed. Briefly, approximately 50 mg of skin was homogenized in 1 ml of TRI Reagent (Sigma, St. Louis, MO, United States) by shaking using 1.4 mm silica beads, then 100 µl of 1-bromo-3-chloropropane (BCP) was added for phase separation. This was followed by precipitation with 500 µl of

isopropanol and posterior washes with 65–75% ethanol. The RNA was then resuspended in RNase-free water and treated with Turbo DNase (Ambion). Samples were then cleaned up using Qiagen RNeasy Mini kit columns and their integrity was checked on Agilent 2200 Bioanalyzer (Agilent Technologies, United States). Thereafter, the Illumina Truseq mRNA stranded RNA-Seq Library Prep Kit protocol was followed directly. Libraries were checked for quality and quantified using the Bioanalyzer 2100 (Agilent), before being sequenced on three lanes of the Illumina HiSeq 4000 instrument using 75 base paired-end sequencing at Edinburgh Genomics, United Kingdom. Raw reads have been deposited in NCBI's Sequence Read Archive (SRA) under Accession No. SRP100978.

Read Mapping

The quality of the sequencing output was assessed using FastQC v.0.11.5.¹ Quality filtering and removal of residual adaptor sequences was conducted on read pairs using Trimmomatic v.0.32 (Bolger et al., 2004). Specifically, Illumina specific adaptors were clipped from the reads, leading, and trailing bases with a Phred score less than 20 were removed and the read trimmed if the sliding window average Phred score over four bases was less than 20. Only reads where both pairs were longer than 36 bp post-filtering were retained. Filtered reads were mapped to the most recent Atlantic salmon genome assembly (ICSASG_v2; GenBank Accession No. GCF_000233375.1; Lien et al., 2016) using STAR v.2.5.2b (Dobin et al., 2013), the maximum number of mismatches for each read pair was set to 10% of trimmed read length, and minimum and maximum intron lengths were set to 20 bases and 1 Mb, respectively. Uniquely mapped paired-reads were counted and assigned to genes (NCBI *S. salar* Annotation Release 100) using FeatureCounts (Liao et al., 2014), included in the SourceForge Subread package v.1.5.0. Only reads with both ends mapped to the same gene were considered in downstream analyses.

Differential Expression

Differential expression analyses and gene functional and pathway enrichment analyses were performed using R v.3.3.1 (R Core Team, 2014). Gene count data were used to estimate differential gene expression using the Bioconductor package DESeq2 v.3.4 (Love et al., 2014). Briefly, size factors were calculated for each sample using the median of ratios method and count data was normalized to account for differences in library depth, next gene-wise dispersion estimates were fitted to the mean intensity using a parametric model and shrunk toward the expected dispersion values, finally a negative binomial model was fitted for each gene and the significance of the coefficients was assessed using the Wald test. The Benjamini–Hochberg false discovery rate (FDR) multiple test correction was applied, and transcripts with FDR < 0.05 and absolute log₂ fold change values (FC) > 0.5 were considered differentially expressed genes. Hierarchical clustering and principal component analyses were performed to visually identify outlier samples that did not cluster close to other samples in the same category (lice attachment site or healthy skin), which

¹<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

were then removed from the analyses as sampling errors could not be discounted. PCA plots were created using the R package *factoextra*.²

Pathway Enrichment

Gene Ontology (GO) enrichment analyses were performed using Blast2GO v.4.1 (Conesa et al., 2005). Briefly, genes showing >10 reads in >90% of the samples were annotated against the manually curated protein database Swiss-Prot (Bairoch et al., 2004) and GO terms were assigned to them using Blast2GO. GO enrichment for specific genes lists was tested against the whole set of expressed genes using Fisher's exact test. GO terms with ≥ 5 DE genes assigned and showing a Benjamini–Hochberg FDR corrected p -value < 0.05 were considered enriched. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using KOBAS v3.0.3 (Wu et al., 2006). Briefly, genes showing >10 reads in >90% of the samples were annotated against KEGG protein database (Kanehisa and Goto, 2000) to determine KEGG Orthology. KEGG enrichment for specific gene lists was tested by comparison to the whole set of expressed genes using Fisher's exact test. KEGG pathways with ≥ 5 differentially expressed (DE) genes assigned and showing a Benjamini–Hochberg FDR corrected p -value < 0.05 were considered enriched.

RESULTS AND DISCUSSION

Disease Challenge

A total of 2,632 fish belonging to 105 families from a commercial breeding program were challenged with *C. rogercresseyi* copepods, and euthanized for sampling 8 days post-challenge. Average lice burden per fish was 38 ± 16 , and the estimated heritability of sea lice load was 0.28 ± 0.04 (unpublished results), therefore the differences in sea lice counts between fish has a genetic component. Fish were selected for RNA sequencing based on the traits of resistance, measured as number and concentration of lice per fish, and weight and length gain since the start of the challenge, which may reflect the ability of the fish to cope with the infestation. The selected fish allowed for 8 vs. 8 comparisons between family-matched fish showing differential resistance (26.2 ± 5.5 vs. 54.9 ± 13.5 sea lice per fish) and differential growth during infestation (7.0 ± 4.3 vs. 28.8 ± 12.3 weight gain percentage). A total of 42 samples (21 fish, skin from sites of louse attachment and healthy skin) were sequenced, resulting in an average of $\sim 27.9 \pm 2.7$ million reads per sample. After trimming, these were aligned against the salmon reference genome (ICSASG_v2; GenBank Accession No. GCF_000233375.1; Lien et al., 2016) and levels of gene expression were estimated according to the official salmon genome annotation (NCBI *S. salar* Annotation Release 100). An average of 19 M trimmed reads per sample were assigned to genes and used for downstream analyses of gene expression. All raw sequence data is available in NCBI's SRA under BioProject Accession No. SRP100978, and may be a useful contribution

to the functional annotation of all salmonid genomes initiative (FAASG; Macqueen et al., 2017).

Louse Attachment Sites Versus Healthy Skin

Principal component analysis of gene expression (Figure 1) revealed a relatively clear cluster of healthy skin samples, while lice-attachment samples were more scattered, probably reflecting variation in the individual response to sea lice. Differential expression between healthy and louse attachment sites resulted in 4,355 DE genes (Supplementary File S2), with a higher number of up-regulated (more expressed in attachment sites) than down-regulated genes ($n = 3,114$ vs. $n = 1,241$). Among these DE genes were well-known components of the innate immune response such as interleukins, interferon response factors and complement components (Figure 2A). GO term and KEGG pathway analyses (Supplementary File S2) revealed a clear enrichment of immune pathways and functions among the up-regulated genes (Figure 2B), highlighting a localized immune response strongly related to cytokine activity. A similar scenario has been observed in other salmonids such as coho salmon where resistance to sea lice has been associated with early inflammation in skin and head kidney, which results in epithelial hyperplasia and often parasite encapsulation and removal of the sea lice within 2 weeks (Johnson and Albright, 1992; Fast et al., 2002). In pink salmon, an early and high expression of pro-inflammatory genes (IL-8, TNF α -1, and IL-1 β) has been suggested as a mechanism of rapid louse rejection (Fast et al., 2007). The classical complement pathway has also been linked to resistance of host fish to parasitic copepod infection (Fast, 2014). The results presented here indicate that despite a marked up-regulation of the local inflammatory response and complement pathway in Atlantic salmon, in part resembling the response of coho salmon or pink salmon, this does not seem to be sufficient to successfully respond to the louse attachment and feeding.

In addition to the expected innate immune response observed above, cell division related processes were also clearly up-regulated at louse attachment sites, and well-characterized genes involved in tissue repair such as fibroblast growth factor-binding protein 1 and Epigen showed significant differences between lice attachment sites and healthy skin ($FC > 3$). Several genes related to the cell matrix and cell adhesion also had higher expression at attachment sites (i.e., cadherin-13, integrin alpha-2, desmoplakin, or various keratin and collagen genes). Cell proliferation is the main response to skin wounds in fish (Iger and Abraham, 1990), and these results are consistent with those previously found in the early response to *L. salmonis* (Skugor et al., 2008). Several mucins were also found to have higher expression at attachment sites, pointing toward increased mucus production and secretion, which can also be a typical response to wounding in fish (Fast, 2014).

Resistance

Resistance, measured as number of sea lice per fish, was evaluated using two different approaches: correlation between

²<http://www.sthda.com/english/rpks/factoextra/>

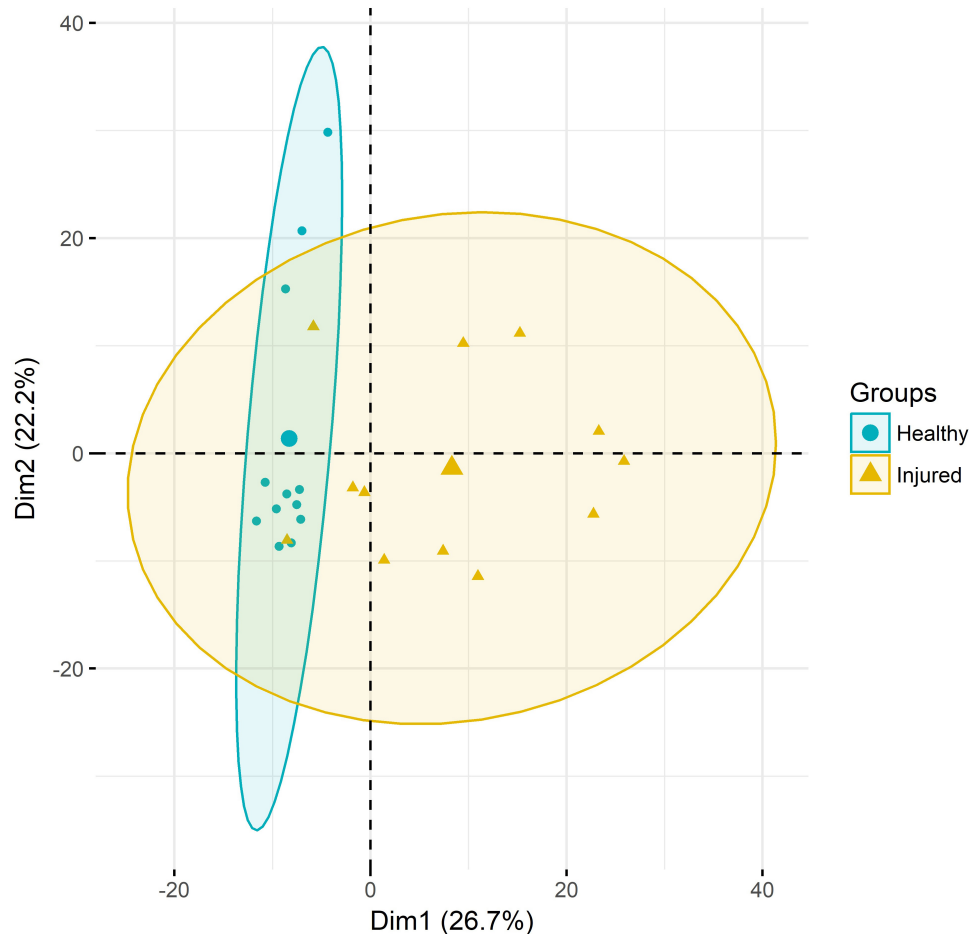


FIGURE 1 | Principal component analyses. RNA-Seq samples clustered according to their gene expression. Ellipses represent 95% confidence intervals.

gene expression and sea lice loads, and differential expression between family-matched fish showing high and low sea lice loads.

Correlation Between Gene Expression and Sea Lice Loads

We studied the correlation of gene expression and sea lice counts in healthy skin and sea lice attachment sites. Genes showing $r > |0.75|$ with sea lice counts were considered of interest (Supplementary File S3).

The expression levels of five immune receptors in healthy skin were positively correlated with sea lice loads (**Table 1**). Macrophage mannose receptor 1 (MRC1) shows the highest positive correlation with number of sea lice ($r = 0.87$), and also the highest expression difference between louse attachment vs. healthy skin (FC = 4.79). MRC1 is a c-type lectin receptor, expressed in macrophages, dendritic cells, and skin in humans. MRC1 plays a role both in innate and adaptive immunity and also acts as a recognition receptor for different pathogens such as bacteria, virus, or fungi (East and Isacke, 2002). C-type lectin receptor A ($r = 0.81$; FC = -1.34) is another lectin receptor involved in antigen recognition and immune response (Geiktenbeek and Gringhuis, 2009). Lectins such as MRC1

and CLEC4E have been found to be induced by glucosinolate-enriched feeds in Atlantic salmon, which also reduced lice counts between 17 and 25% (Holm et al., 2016), and are also up-regulated in response to sea lice in the more resistant pink salmon species (Sutherland et al., 2014). Lectins have been reported to activate the immune system in response to parasites in several different species (Vázquez-Mendoza et al., 2013; Hoving et al., 2014), therefore modulation of these genes represents a possible route to enhance Atlantic immune responses to sea lice. Two immune receptors were negatively correlated with number of sea lice, CD97 ($r = -0.84$) and suppressor of cytokine signaling 5 (SOCS5; $r = -0.76$). CD97 regulates cytokine production and T-cell activation and proliferation (Capasso et al., 2006; Abbott et al., 2007); while SOCS5 is part of the cytokine-mediated signaling pathway, and acts as a negative regulator of inflammatory response and other immune-related pathways (Seki et al., 2002). Since, it was not possible to take skin samples prior to infection, it is difficult to distinguish between cause and effect; i.e., it is plausible that the negative correlation of these genes with number of sea lice is simply indicating that the immune system of the host responds proportionally to the degree of lice infestation. Nonetheless, the data support a major role for

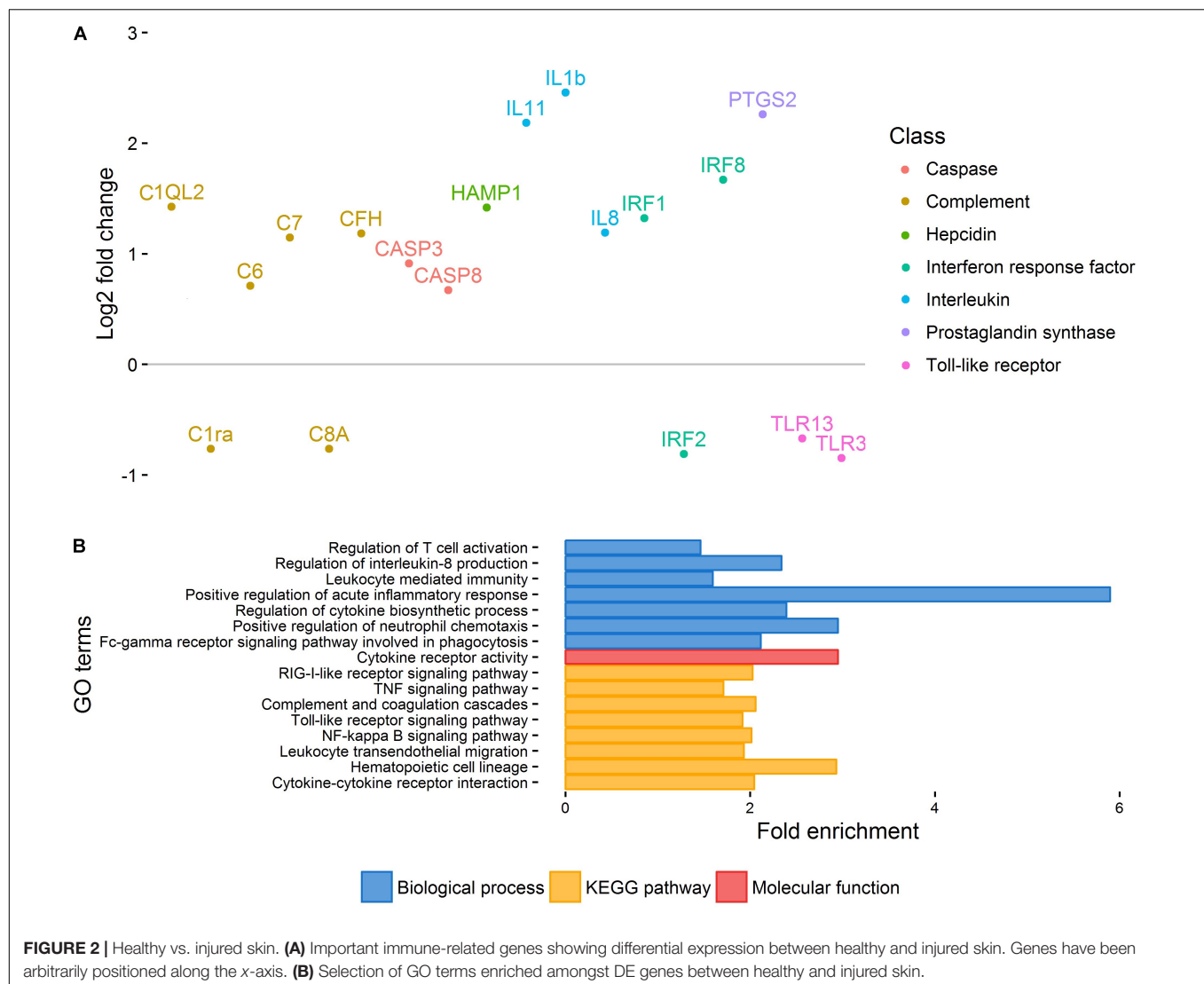


TABLE 1 | Immune receptors showing correlation with sea lice counts.

| Gene | Full name | Corr. (r) | GO terms |
|--------|---|-----------|---|
| MRC1 | Macrophage mannose receptor 1 | 0.87 | Cellular response to interferon gamma Cellular response to interleukin-4 |
| CLEC4E | C-type lectin receptor A | 0.81 | Innate immune response Positive regulation of cytokine secretion |
| CR2 | Complement receptor type 2 | 0.78 | Innate immune response Complement activation, classical pathway |
| LIFR | Leukemia inhibitory factor receptor | 0.78 | Cytokine-mediated signaling pathway |
| CD28 | T-cell-specific surface glycoprotein CD28 | 0.77 | Positive regulation of inflammatory response Response to antigenic stimulus Cytokine biosynthetic process |
| SOCS5 | Suppressor of cytokine signaling 5 | -0.76 | Cytokine-mediated signaling pathway Negative regulation of inflammatory response |
| CD97 | CD97 antigen | -0.84 | Inflammatory response |

these genes in the host response to sea lice, and the differences in lice count between fish has a genetic component, to which these genes may contribute.

Amongst genes without a (well-known) immune function, there was an association between SUMO1 ($r = 0.76$) and SUMO3 ($r = -0.91$) expression and sea lice loads. Small ubiquitin-like modifier (SUMO) proteins are small proteins similar to ubiquitins that are covalently attached to other proteins to modify

their function. According to the gene expression data, SUMO1 seems to be preferred over SUMO3 in salmon upon sea lice infestation. Although post-translational modifications have been barely explored in fish, in mice SUMOylation has been shown to be involved in modulation of host innate immune response to pathogens (Decque et al., 2016). SUMOylation is also a very active field of research in plants, where SUMO is known to be involved in many important processes such as plant response to

environmental stresses, including pathogens (Park et al., 2011). It would be interesting to further study the role of SUMO in modulating Atlantic salmon responses to sea lice.

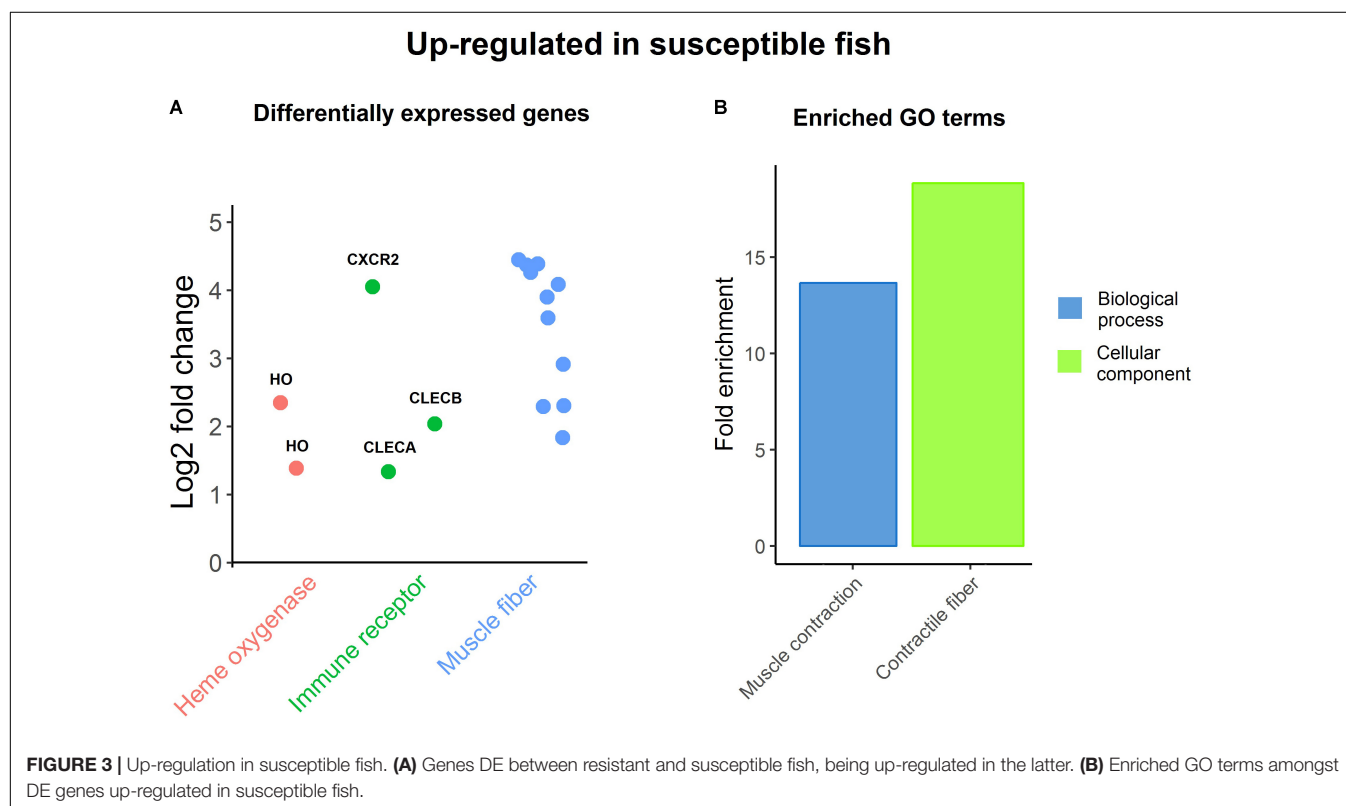
The results were markedly different in lice-attachment sites (Supplementary File S3), and congruent with differential expression between lice-attachment and healthy skin, with inflammatory genes such as toll-like receptor 12 or caspase 3 showing high correlations with sea lice loads. Similarly, one of the *sox9* paralogs (*sox9a*) was also highly correlated with lice loads. *Sox9* has a pro-proliferation function in human epidermal keratinocytes (Shi et al., 2013), and therefore this transcription factor is probably promoting wound healing in sea lice attachment sites. Finally, the gene *hepcidin-1* is also correlated with sea lice counts in lice-attachment sites. *Hepcidin* is a regulator of iron metabolism, which as mentioned in the introduction has been associated with response to *C. rogercresseyi* (Valenzuela-Muñoz et al., 2017).

Differential Expression Between High and Low Sea Lice Loads

The samples for RNA sequencing were chosen to enable 8 vs. 8 comparison between family-matched fish (three families with two fish per group, two families with one fish) with high and low values for resistance (26.2 ± 5.5 vs. 54.9 ± 13.5 sea lice per fish). There were 43 genes significantly differentially expressed between resistant and susceptible fish (Supplementary File S4). All but one were from comparison of healthy skin samples between the two groups, which seems to suggest that the differences in resistance are systemic rather than local to louse attachment sites. The

susceptible group had higher expression levels for genes involved in muscle contraction like troponins and myosins, which was also highlighted by GO enrichment analyses (Figure 3). Myosins and troponins have previously been identified as genes that respond to sea lice attachment in salmon skin (Holm et al., 2015). Further, *Caligus* infection is known to induce increased enzyme activity in muscle tissue (Vargas-Chacoff et al., 2017), and behavioral changes in the fish such as flashing and jumping are associated with ectoparasite removal (Furevik et al., 1993; Magnhagen et al., 2008). It has been recently reported that inactivity or reduced swimming activity contribute to resistance to sea lice (Bui, 2017), so it is possible that the high lice counts of susceptible fish in this study are due to higher activity levels with associated expression of muscle contraction related genes. In turn, high lice burden can provoke behavioral responses increasing fish activity, which results in the up-regulation of muscle genes, increasing the expression differences between resistant-passive-low lice fish and susceptible-active-high lice fish.

Two heme oxygenase genes, encoding enzymes, which catalyze the degradation of heme, also had higher expression levels in susceptible samples (Figure 3), which is consistent with the positive correlation with lice loads of the iron-sequestration gene *hepcidin*. These genes have been previously shown to be up-regulated in response to *Caligus* infection (Valenzuela-Muñoz and Gallardo-Escárate, 2017). Importantly, iron availability was found to be reduced in the highly resistant species pink salmon infected with *L. Salmonis* (Sutherland et al., 2014), and hematocrit and anemia were also found to be reduced in chum salmon (*Oncorhynchus keta*) in response to



sea lice (Jones et al., 2007). It is therefore plausible that the more effective reduction of iron availability in Atlantic salmon (perhaps behaving more similarly to the resistant pink salmon) might be related to increased resistance to sea lice.

Finally, three immune receptors showed higher expression in susceptible samples (Figure 3); C-X-C chemokine receptor type 2 is a receptor for IL-8, its binding causes activation of neutrophils; while C type lectin receptors A (also found to be positively correlated with sea lice counts in healthy skin) and B are leptin receptors with an important role in pathogen recognition and immunity (Geiktenbeek and Gringhuis, 2009), as previously discussed. While it is clear that resistance and host response to sea lice is multifactorial in nature, these genes related to muscle contraction, iron availability and immunity may be targets for functional validation in future studies, and for cross-referencing with genome-wide association analyses to identify candidate causative genes and variants.

Growth During Infestation

Differences in weight gain percentage from the start to the end of the trial were also investigated. Weight gain during infestation did not show any significant correlation with initial weight ($r = -0.27$, $p = 0.10$), sea lice counts ($r = 0.12$, $p = 0.45$) or sea lice density ($r = 0.19$, $p = 0.24$) in our dataset, and the means for these three traits are not significantly different between our groups showing differential growth during infestation (t -test p -values > 0.35). Family-matched fish (8 vs. 8; three families with two fish per group, two families with one fish) with differential weight gain during infestation (7.0 ± 4.3 vs. 28.8 ± 12.3 weight gain percentage) were compared. A total of 24 and 1 genes were found differentially expressed between fish showing high and low weight gains in healthy and sea louse attachment site samples, respectively (Supplementary File S5). The gene differentially expressed in injured skin, solute carrier family 15 member 1 (SLC15A1), also showed the lowest p -value and highest FC in healthy skin (FC = 3.38, $p = 0.003$). The SLC15A1 protein is a membrane transporter that mediates the uptake of dipeptides and tripeptides, in humans this gene is expressed in the intestinal epithelium and plays a major role in protein absorption (Adibi, 1997). Another interesting DE gene is myogenic regulatory factor 6 (MYF6; FC = 0.72, $p = 0.04$). Myogenic regulatory factors are transcription factors that regulate muscle development (Perry and Rudnick, 2000); in *Senegalese sole* decreased expression of these factors was observed in fast muscle when fed with a high-lipid content diet, which caused reduced growth (Campos et al., 2010). While skin is unlikely to be a highly suitable tissue to study genes underlying fish growth during sea lice infestation, both myogenic factors and increased nutrient absorption, and specifically MYF6 and SLC15A1, are good candidates to better understand growth impairment differences under sea lice infestation.

CONCLUSION

The results of this study highlight that the early gene expression response of Atlantic salmon to sea lice involves up-regulation of

many different components of the immune system (inflammatory response, cytokine production, TNF and NF-kappa B signaling and complement activation) along with tissue repair activation. The comparison of resistant vs. susceptible animals highlighted enrichment of pathways related to fish activity, iron availability and receptors modulating pathogen recognition and immune response. Overall, this study contributes to an improved understanding of Atlantic salmon early response to sea lice in skin, and into the gene expression profiles underpinning genetic resistance to sea lice in salmon. The identified pathways and genes may be targets for future studies aimed at development of new treatments, vaccines, or prevention strategies. The data can also be cross-referenced with high power genome-wide association studies to help prioritize putative causative genes and variants that have potential to improve genomic selection programs for genetic improvement of resistance to this industry's most serious disease.

DATA AVAILABILITY

The raw reads generated for this study have been deposited in NCBI's Sequence Read Archive (SRA) under Accession No. SRP100978.

ETHICS STATEMENT

The lice challenge experiments were performed under local and national regulatory systems and were approved by the Animal Bioethics Committee (ABC) of the Faculty of Veterinary and Animal Sciences of the University of Chile (Santiago, Chile), Certificate No. 01-2016, which based its decision on the Council for International Organizations of Medical Sciences (CIOMS) standards, in accordance with the Chilean standard NCh-324-2011.

AUTHOR CONTRIBUTIONS

RH, JY, and DR were responsible for the concept and design of this work and drafted the manuscript. AB managed the collection of the samples. AG performed the molecular biology experiments. DR performed bioinformatic and statistical analyses. All authors read and approved the final manuscript.

FUNDING

This work was supported by an RCUK-CONICYT (Grant No. BB/N024044/1), Institute Strategic Funding Grants to The Roslin Institute (Grant Nos. BBS/E/D/20002172, BBS/E/D/30002275, and BBS/E/D/10002070), Edinburgh Genomics was partly supported through core grants from NERC (Grant No. R8/H10/56), MRC (Grant No. MR/K001744/1), and BBSRC (Grant No. BB/J004243/1). DR was supported by a Newton International Fellowship of the Royal Society (Grant No. NF160037).

ACKNOWLEDGMENTS

The authors would like to thank the contribution of Aquainnovo and Salmones Chaicas for providing the biological material and phenotypic records of the experimental challenges.

REFERENCES

- Aaen, S. M., Helgesen, K. O., Bakke, M. K., Kaur, K., and Horsberg, T. E. (2015). Drug resistance in sea lice: a threat to salmonid aquaculture. *Trends Parasitol.* 31, 72–81. doi: 10.1016/j.pt.2014.12.006
- Abbott, R. J., Spendlove, I., Roversi, P., Fitzgibbon, H., Knott, V., Teriete, P., et al. (2007). Structural and functional characterization of a novel T cell receptor co-regulatory protein complex. CD97-CD55. *J. Biol. Chem.* 282, 22023–22032. doi: 10.1074/jbc.M702588200
- Adibi, S. A. (1997). The oligopeptide transporter (Pept-1) in human intestine: biology and function. *Gastroenterology* 113, 332–340. doi: 10.1016/S0016-5085(97)70112-4
- Bairoch, A., Boeckmann, B., Ferro, S., and Gasteiger, E. (2004). Swiss-Prot: juggling between evolution and stability. *Brief. Bioinform.* 5, 39–55. doi: 10.1093/bib/5.1.39
- Bolger, A. M., Lohse, M., and Usadel, B. (2004). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Braden, L. M., Barker, D. E., Koop, B. F., and Jones, S. R. M. (2012). Comparative defense-associated responses in salmon skin elicited by the ectoparasite *Lepeophtheirus salmonis*. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 7, 100–109. doi: 10.1016/j.cbpd.2011.12.002
- Braden, L. M., Koop, B. F., and Jones, S. R. M. (2015). Signatures of resistance to *Lepeophtheirus salmonis* include a TH2-type response at the louse-salmon interface. *Dev. Comp. Immunol.* 48, 178–191. doi: 10.1016/j.dci.2014.09.015
- Bui, S. (2017). *The Behavioural Resistance and Response of Atlantic Salmon to the Ectoparasite Lepeophtheirus salmonis*. Ph.D. thesis, University of Melbourne, Parkville VIC.
- Campos, C., Valente, L. M. P., Borges, P., Bizuayehu, T., and Fernandes, J. M. O. (2010). Dietary lipid levels have a remarkable impact on the expression of growth-related genes in Senegalese sole (*Solea senegalensis* Kaup). *J. Exp. Biol.* 213, 200–209. doi: 10.1242/jeb.033126
- Capasso, M., Durrant, L. G., Stacey, M., Gordon, S., Ramage, J., and Spendlove, I. (2006). Costimulation via CD55 on human CD4+ T cells mediated by CD97. *J. Immunol.* 177, 1070–1077. doi: 10.4049/jimmunol.177.2.1070
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., and Yáñez, J. M. (2017). 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. doi: 10.1186/s12711-017-0291-8
- Correa, K., Lhorente, J. P., Bassini, L., López, M. E., Di Genova, A., Maass, A., et al. (2016). 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. doi: 10.1016/j.aquaculture.2016.04.008
- Costello, M. J. (2009). The global economic cost of sea lice to the salmonid farming industry. *J. Fish Dis.* 32, 115–118. doi: 10.1111/j.1365-2761.2008.01011.x
- Decque, A., Joffre, O., Magalhaes, J. G., Cossec, J. C., Blecher-Gonen, R., Lapaguet, P., et al. (2016). Sumoylation coordinates the repression of inflammatory and anti-viral gene-expression programs during innate sensing. *Nat. Immunol.* 17, 140–149. doi: 10.1038/ni.3342
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- East, L., and Isacke, C. M. (2002). The mannose receptor family. *Biochim. Biophys. Acta* 1572, 364–386. doi: 10.1016/S0304-4165(02)00319-7
- Edvardsen, R. B., Leininger, S., Kleppe, L., Skafnesmo, K. O., and Wargelius, A. (2014). Targeted mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. *PLoS One* 9:e108622. doi: 10.1371/journal.pone.0108622
- Fast, M. D. (2014). Fish immune responses to parasitic copepod (namely sea lice) infection. *Dev. Comp. Immunol.* 43, 300–312. doi: 10.1016/j.dci.2013.08.019
- Fast, M. D., Johnson, S. C., and Jones, S. R. M. (2007). Differential susceptibility and the response of pink (*Oncorhynchus gorbuscha*) and chum (*O. keta*) salmon juveniles to infection with *Lepeophtheirus salmonis*. *Dis. Aquat. Organ.* 75, 229–238.
- Fast, M. D., Ross, N. W., Mustafa, A., Sims, D. E., Johnson, S. C., Conboy, G. A., et al. (2002). Susceptibility of rainbow trout *Oncorhynchus mykiss*, Atlantic salmon *Salmo salar* and coho salmon *Oncorhynchus kisutch* to experimental infection with sea lice *Lepeophtheirus salmonis*. *Dis. Aquat. Org.* 52, 57–68. doi: 10.3354/dao052057
- Food and Agriculture Organization of the United Nations (2016). *The State of World Fisheries and Aquaculture 2016*. Rome: Contributing to food security and nutrition for all.
- Food and Agriculture Organization of the United Nations (2017). *GLOBEFISH Highlights – Special Issue 2017*. Rome: Food and Agriculture Organization of the United Nations, 24. doi: 10.1098/rspb.2011.2210
- Frazer, L. N., Morton, A., and Krkosek, M. (2012). Critical thresholds in sea lice epidemics: evidence, sensitivity and subcritical estimation. *Proc. Biol. Sci.* 279, 1950–1958. doi: 10.1098/rspb.2011.2210
- Furevik, D. M., Bjørndal, A., Huse, I., and Fernö, A. (1993). Surface activity of Atlantic salmon (*Salmo salar* L.) in net pens. *Aquaculture* 110, 119–128. doi: 10.1038/nri2569
- Geiktenbeek, T. B., and Gringhuis, S. I. (2009). Signalling through C-type lectin receptors: shaping immune responses. *Nat. Rev. Immunol.* 9, 465–479. doi: 10.1038/nri2569
- Gharbi, K., Matthews, L., Bron, J., Roberts, R., Tinch, A., and Stear, M. (2015). The control of sea lice in Atlantic salmon by selective breeding. *J. R. Soc. Interface* 12:0574. doi: 10.1098/rsif.2015.0574
- Holm, H., Santi, N., Kjøglum, S., Perisic, N., Skugor, S., and Evensen, Ø. (2015). Difference in skin immune responses to infection with salmon louse (*Lepeophtheirus salmonis*) in Atlantic salmon (*Salmo salar* L.) of families selected for resistance and susceptibility. *Fish Shellfish Immunol.* 42, 384–394. doi: 10.1016/j.fsi.2014.10.038
- Holm, H. J., Wadsworth, S., Bjelland, A. K., Krasnov, A., Evensen, Ø., and Skugor, S. (2016). Dietary phytochemicals modulate skin gene expression profiles and result in reduced lice counts after experimental infection in Atlantic salmon. *Parasit. Vectors* 9:271. doi: 10.1186/s13071-016-1537-y
- Hoving, J. C., Wilson, G. J., and Brown, G. D. (2014). Signalling C-type lectin receptors, microbial recognition and immunity. *Cell Microbiol.* 16, 185–194. doi: 10.1111/cmi.12249
- Iger, Y., and Abraham, M. (1990). The process of skin healing in experimentally wounded carp. *J. Fish Biol.* 36, 421–437. doi: 10.1016/j.aquaculture.2013.12.033
- Imsland, A., Reynolds, P., Eliassen, G., Hangstad, T. A., Foss, A., Vikingstad, E., et al. (2014). The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeophtheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar*). *Aquaculture* 42, 19–23. doi: 10.1186/s13071-016-1537-y
- Jodaa, H. H., Wadsworth, S., Bjelland, A. K., Krasnov, A., Evensen, Ø., and Skugor, S. (2016). Dietary phytochemicals modulate skin gene expression profiles and result in reduced lice counts after experimental infection in Atlantic salmon. *Parasit. Vectors* 9:271. doi: 10.1186/s13071-016-1537-y
- Johnson, S. C., and Albright, L. J. (1992). Comparative susceptibility and histopathology of the response of naive Atlantic, Chinook and coho salmon to experimental infection with *Lepeophtheirus salmonis* (Copepoda, Caligidae). *Dis. Aquat. Organ.* 14, 179–193.
- Johnson, S. C., Treasurer, J. W., Bravo, S., Nagasawa, K., and Kabata, Z. (2004). A review of the impact of parasitic copepods on marine aquaculture. *Zool. Stud.* 43, 229–243. doi: 10.3354/dao075229

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00287/full#supplementary-material>

- Jones, S. R. M., Fast, M. D., Johnson, S. C., and Groman, D. B. (2007). Differential rejection of salmon lice by pink and chum salmon: disease consequences and expression of proinflammatory genes. *Dis. Aquat. Org.* 75, 229–238. doi: 10.1093/nar/28.1.27
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1146/annurev-marine-010814-015646
- Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., et al. (2015). Infectious diseases affect marine fisheries and aquaculture economics. *Ann. Rev. Mar. Sci.* 7, 471–496. doi: 10.1146/annurev-marine-010814-015646
- Leclercq, E., Davie, A., and Migaud, H. (2014). Delousing efficiency of farmed ballan wrasse (*Labrus bergylta*) against *Lepeophtheirus salmonis* infecting Atlantic salmon (*Salmo salar*) post-smolts. *Pest. Manag. Sci.* 70, 1274–1282. doi: 10.1002/ps.3692
- Lhorente, J. P., Gallardo, J. A., Villanueva, B., Carabaño, M. J., and Neira, R. (2014). Disease resistance in Atlantic salmon (*Salmo salar*): coinfection of the intracellular bacterial pathogen *Piscirickettsia salmonis* and the sea louse *Caligus rogercresseyi*. *PLoS One* 9:e95397. doi: 10.1371/journal.pone.0095397
- Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533, 200–205. doi: 10.1038/nature17164
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s12864-016-2443-6
- MacLeod, I. M., Bowman, P. J., Vander Jaqt, C. J., Haile-Mariam, M., Kemper, K. E., Chamberlain, A. J., et al. (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* 17:144. doi: 10.1186/s12864-016-2443-6
- Macqueen, D. J., Primmer, C. R., Houston, R. D., Nowak, B. F., Bernatchez, L., Bergseth, S., et al. (2017). Functional annotation of all salmonid genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics* 18:484.
- Magnhagen, C., Braithwaite, V., Forsgren, E., and Kapoor, B. G. (2008). *Fish Behaviour*. Mount Washington, NH: Enfield Science Publishers.
- Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., and Meuwissen, T. H. E. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402
- Palaikostas, C., Ferrareso, S., Franch, R., Houston, R. D., and Bargelloni, L. (2016). Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3* 6, 3693–3700. doi: 10.1534/g3.116.035220
- Park, H. J., Kim, W. Y., Park, H. C., Lee, S. Y., Bohnert, H. J., and Yun, D. J. (2011). SUMO and SUMOylation in plants. *Mol. Cells* 32, 305–316. doi: 10.1007/s10059-011-0122-7
- Perry, R., and Rudnick, M. (2000). Molecular mechanisms regulating myogenic determination and differentiation. *Front. Biosci.* 5:D750–D767.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Seki, Y., Hayashi, K., Matsumoto, A., Tsukada, J., Ransom, J., Naka, T., et al. (2002). Expression of the suppressor of cytokine signaling-5 (SOCS5) negatively regulates IL-4-dependent STAT6 activation and Th2 differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13003–13008. doi: 10.1073/pnas.202477099
- Shi, G., Sohn, K. C., Li, Z., Choi, D. K., Park, Y. M., Kim, J. H., et al. (2013). Expression and functional role of Sox9 in human epidermal keratinocytes. *PLoS One* 8:e54355. doi: 10.1371/journal.pone.0054355
- Skuger, S., Glover, K. A., and Nilsen, F. (2008). Local and systemic gene expression responses of Atlantic salmon (*Salmo salar* L.) to infection with the salmon louse (*Lepeophtheirus salmonis*). *BMC Genomics* 9:498. doi: 10.1186/1471-2164-9-498
- Sutherland, B. J. G., Covello, J. M., Friend, S. E., Poley, J. D., Koczka, K. W., Purcell, S. L., et al. (2017). Host-parasite transcriptomics during immunostimulant-enhanced rejection of salmon lice (*Lepeophtheirus salmonis*) by Atlantic salmon (*Salmo salar*). *FACETS* 2, 477–495. doi: 10.1139/facets-2017-0020
- Sutherland, B. J. G., Koczka, K. W., Yasuike, M., Jantzen, S. G., Yazawa, R., Koop, B. F., et al. (2014). Comparative transcriptomics of Atlantic *Salmo salar*, chum *Oncorhynchus keta* and pink salmon *O. gorbuscha* during infections with salmon lice *Lepeophtheirus salmonis*. *BMC Genomics* 15:200. doi: 10.1186/1471-2164-15-200
- Sutherland, B. J. G., Poley, J. D., Igboeli, O. O., Jantzen, J. R., Fast, M. D., Koop, B. F., et al. (2015). Transcriptomic responses to emamectin benzoate in Pacific and Atlantic Canada salmon lice *Lepeophtheirus salmonis* with differing levels of drug resistance. *Evol. Appl.* 8, 133–148. doi: 10.1111/eva.12237
- Torrisen, O., Jones, S., Asche, F., Guttomsen, A., Skilbrei, O. T., Nilsen, F., et al. (2013). Salmon lice – impact on wild salmonids and salmon aquaculture. *J. Fish Dis.* 36, 171–194. doi: 10.1111/jfd.12061
- Tsai, H. Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48:47. doi: 10.1186/s12711-016-0226-9
- Valenzuela-Muñoz, V., Boltaña, S., and Gallardo-Escárate, C. (2016). Comparative immunity of *Salmo salar* and *Oncorhynchus kisutch* during infestation with the sea louse *Caligus rogercresseyi*: an enrichment transcriptome analysis. *Fish Shellfish Immunol.* 59, 276–287. doi: 10.1016/j.fsi.2016.10.046
- Valenzuela-Muñoz, V., Boltaña, S., and Gallardo-Escárate, C. (2017). Uncovering iron regulation with species-specific transcriptome patterns in Atlantic and coho salmon during a *Caligus rogercresseyi* infestation. *J. Fish Dis.* 40, 1169–1184. doi: 10.1111/jfd.12592
- Valenzuela-Muñoz, V., and Gallardo-Escárate, C. (2017). Iron metabolism modulation in Atlantic salmon infested with the sea lice *Lepeophtheirus salmonis* and *Caligus rogercresseyi*: a matter of nutritional immunity? *Fish Shellfish Immunol.* 60, 97–102. doi: 10.1016/j.fsi.2016.11.045
- Vargas-Chacoff, L., Muñoz, J. L. P., Hawes, C., Oyarzún, R., Pontigo, J. P., Saravia, J., et al. (2017). Ectoparasite *Caligus rogercresseyi* modifies the lactate response in Atlantic salmon (*Salmo salar*) and coho salmon (*Oncorhynchus kisutch*). *Vet. Parasitol.* 243, 6–11. doi: 10.1016/j.vetpar.2017.05.031
- Vázquez-Mendoza, A., Carrero, J. C., and Rodríguez-Sosa, M. (2013). Parasitic infections: a role for C-type lectins receptors. *Biomed. Res. Int.* 2013:456352. doi: 10.1155/2013/456352
- Wargelius, A., Leininger, S., Skafnesmo, K. O., Kleppe, L., Andersson, E., Taranger, G. L., et al. (2016). Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Sci. Rep.* 6:21284. doi: 10.1038/srep21284
- Wu, J., Mao, X., Cai, T., Luo, J., and Wei, L. (2006). KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* 34, W720–W724. doi: 10.1093/nar/gkl167
- Yáñez, J. M., Houston, R. D., and Newman, S. (2014a). Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5:415. doi: 10.3389/fgene.2014.00415
- Yáñez, J. M., Lhorente, J. P., Bassini, L. N., Oyarzún, M., Neira, R., and 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. doi: 10.1016/j.aquaculture.2014.06.026

Conflict of Interest Statement: JY was supported by Aquainnovo S.A.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Robledo, Gutiérrez, Barria, Yáñez and Houston. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



RNA Sequencing (RNA-Seq) Reveals Extremely Low Levels of Reticulocyte-Derived Globin Gene Transcripts in Peripheral Blood From Horses (*Equus caballus*) and Cattle (*Bos taurus*)

OPEN ACCESS

Edited by:

John Anthony Hammond,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Chuanju Dong,
Henan Normal University, China
Luiz Lehmann Coutinho,
Universidade de São Paulo, Brazil

*Correspondence:

David E. MacHugh
david.machugh@ucd.ie

†Present Address:

Nicolas C. Nalpas,
Quantitative Proteomics and
Proteome Centre Tübingen,
Interfaculty Institute for Cell Biology,
University of Tübingen, Tübingen,
Germany
Kevin Rue-Albrecht,
Kennedy Institute of Rheumatology,
University of Oxford, Oxford,
United Kingdom

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 03 April 2018

Accepted: 09 July 2018

Published: 14 August 2018

Citation:

Correia CN, McLoughlin KE,
Nalpas NC, Magee DA, Browne JA,
Rue-Albrecht K, Gordon SV and
MacHugh DE (2018) RNA Sequencing
(RNA-Seq) Reveals Extremely Low
Levels of Reticulocyte-Derived Globin
Gene Transcripts in Peripheral Blood
From Horses (*Equus caballus*) and
Cattle (*Bos taurus*).
Front. Genet. 9:278.
doi: 10.3389/fgene.2018.00278

Carolina N. Correia¹, Kirsten E. McLoughlin¹, Nicolas C. Nalpas^{1†}, David A. Magee¹,
John A. Browne¹, Kevin Rue-Albrecht^{1†}, Stephen V. Gordon^{2,3} and David E. MacHugh^{1,3*}

¹ Animal Genomics Laboratory, UCD School of Agriculture and Food Science, UCD College of Health and Agricultural Sciences, University College Dublin, Dublin, Ireland, ² UCD School of Veterinary Medicine, UCD College of Health and Agricultural Sciences, University College Dublin, Dublin, Ireland, ³ UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland

RNA-seq has emerged as an important technology for measuring gene expression in peripheral blood samples collected from humans and other vertebrate species. In particular, transcriptomics analyses of whole blood can be used to study immunobiology and develop novel biomarkers of infectious disease. However, an obstacle to these methods in many mammalian species is the presence of reticulocyte-derived globin mRNAs in large quantities, which can complicate RNA-seq library sequencing and impede detection of other mRNA transcripts. A range of supplementary procedures for targeted depletion of globin transcripts have, therefore, been developed to alleviate this problem. Here, we use comparative analyses of RNA-seq data sets generated from human, porcine, equine, and bovine peripheral blood to systematically assess the impact of globin mRNA on routine transcriptome profiling of whole blood in cattle and horses. The results of these analyses demonstrate that total RNA isolated from equine and bovine peripheral blood contains very low levels of globin mRNA transcripts, thereby negating the need for globin depletion and greatly simplifying blood-based transcriptomic studies in these two domestic species.

Keywords: blood, cattle, globin, horses, pigs, reticulocyte, RNA-seq, transcriptome

INTRODUCTION

It is increasingly recognised that new technological approaches are urgently required for infectious disease diagnosis, surveillance, and management in burgeoning domestic animal populations as livestock production intensifies across the globe (Thornton, 2010; Nabarro and Wannous, 2014; Animal Task Force, 2016). In this regard, new strategies have emerged that leverage peripheral blood gene expression to study host immunobiology and to identify panels of RNA transcript biomarkers that can be used as specific biosignatures of infection by particular pathogens for

both animal and human infectious disease (Ramilo and Mejias, 2009; Mejias and Ramilo, 2014; Chaussabel, 2015; Ko et al., 2015; Holcomb et al., 2017). For example, we and others have applied this approach to bovine tuberculosis (BTB) caused by infection with *Mycobacterium bovis* (Meade et al., 2007; Killick et al., 2011; Blanco et al., 2012; Churbanov and Milligan, 2012; McLoughlin et al., 2014; Cheng et al., 2015). It is also important to note that peripheral blood transcriptomics using technologies such as microarrays or RNA-sequencing (RNA-seq) can be used to monitor changes in the physiological status of domestic animals due to reproductive status, diet and nutrition or stress (O'Loughlin et al., 2012; Takahashi et al., 2012; Song et al., 2013; Kolli et al., 2014; Shen et al., 2014; de Greeff et al., 2016; Elgendy et al., 2016; Jégou et al., 2016).

During the last 15 years, a major hindrance to whole blood transcriptomics studies has emerged, which is the presence of large quantities of globin mRNA transcripts in peripheral blood from many mammalian species (Wu et al., 2003; Fan and Hegde, 2005; Liu et al., 2006). This is a consequence of abundant α globin and β globin mRNA transcripts in circulating reticulocytes, which in humans, may account for more than 95% of the total cellular mRNA content in these immature erythrocytes (Debey et al., 2004). Reticulocytes, in turn, account for 1–4% of the erythrocytes in healthy adult humans, which corresponds to between 5×10^7 and 2×10^8 cells per ml compared to 7×10^6 cells per ml for leukocytes (Greer et al., 2013). Hence, globin transcripts can account for a substantial proportion of total detectable mRNAs in peripheral blood samples collected from humans and many other mammals (Bruder et al., 2010; Winn et al., 2010; Schwochow et al., 2012; Choi et al., 2014; Shin et al., 2014; Bowyer et al., 2015; Huang et al., 2016; Morey et al., 2016). In particular, for humans, more than 70% of peripheral blood mRNA transcripts are derived from the haemoglobin subunit alpha 1, subunit alpha 2 and subunit beta genes (*HBA1*, *HBA2*, and *HBB*) (Wu et al., 2003; Field et al., 2007; Mastrokolias et al., 2012).

The emergence of massively parallel transcriptome profiling for clinical applications in human peripheral blood—initially with gene expression microarrays, but more recently using RNA-seq—has prompted development of methods for the systematic reduction of globin mRNAs in total RNA samples purified from peripheral blood samples, including: oligonucleotides that bind to globin mRNA molecules with subsequent digestion of the RNA strand of the RNA:DNA hybrid (Wu et al., 2003); peptide nucleic acid (PNA) oligonucleotides that are complementary to globin mRNAs and block reverse transcription of these targets (Liu et al., 2006); the GLOBINclear™ system, which uses biotinylated oligonucleotides that hybridise with globin transcripts followed by capture and separation using streptavidin-coated magnetic beads (Field et al., 2007); and the recently introduced GlobinLock method that uses a pair of modified oligonucleotides complementary to the 3' portion of globin transcripts and that block enzymatic extension (Křutškov et al., 2016).

In the present study we use RNA-seq data generated from globin-depleted and non-depleted total RNA purified from human and porcine peripheral blood, in conjunction with

non-depleted total RNA isolated from equine and bovine peripheral blood, for a comparative investigation of the impact of reticulocyte-derived globin mRNA transcripts on routine transcriptome profiling of blood in domestic cattle and horses. The primary objective of the present study to test the hypothesis that both cattle and horses exhibit significantly lower quantities of haemoglobin gene transcripts compared to humans and pigs.

MATERIALS AND METHODS

Data Sources

RNA-seq data sets from human peripheral whole blood samples used for assessment of globin depletion and with parallel non-depleted controls (Shin et al., 2014) were obtained from the NCBI Gene Expression Omnibus (GEO) database (accession number GSE53655). A comparable RNA-seq data set from globin-depleted and non-depleted porcine peripheral whole blood was obtained directly from the study authors (Choi et al., 2014). A published RNA-seq data set (Ropka-Molik et al., 2017) from equine non-depleted peripheral whole blood was obtained from the NCBI GEO database (accession number GSE83404). Finally, bovine RNA-seq data from peripheral whole blood were generated by us as described below and can be obtained from the European Nucleotide Archive (ENA) database (PRJEB27764). A summary overview of the methodology used for the current study is shown in Figure 1.

Human, Porcine and Equine Sample Collection, Globin Depletion, and RNA-Seq Libraries

Detailed information concerning ethics approval, sample collection, total RNA extraction, and RNA-seq library preparation and sequencing for the human, porcine, and equine data sets is provided in the original publications (Choi et al., 2014; Shin et al., 2014; Ropka-Molik et al., 2017). **Supplementary Table 1** provides summary information on the human, porcine and equine samples and RNA-seq libraries.

In brief, for the human samples, peripheral blood from six healthy subjects (three females and three males) was collected into PAXgene blood RNA tubes (PreAnalytiX/Qiagen Ltd., Manchester, UK). Total RNA, including small RNAs, was purified from the collected blood samples using the PAXgene Blood miRNA Kit (PreAnalytiX/Qiagen Ltd.) as described by Shin et al. (2014). Human *HBA1*, *HBA2* and *HBB* mRNA transcripts were depleted from a subset of the total RNA samples using the GLOBINclear kit (Invitrogen™/Thermo Fisher Scientific, Loughborough, UK). RNA-seq data was then generated using 24 paired-end (PE) RNA-seq libraries (12 undepleted and 12 globin-depleted) generated from the six biological replicates and six identical technical replicates created from pooled total RNA across all six donor samples. The multiplexing and sequencing was then performed such that data for the 12 samples in each treatment group (undepleted and globin depleted) was generated from two separate lanes of a single flow cell twice, for a total of four sequencing lanes (Shin et al., 2014).

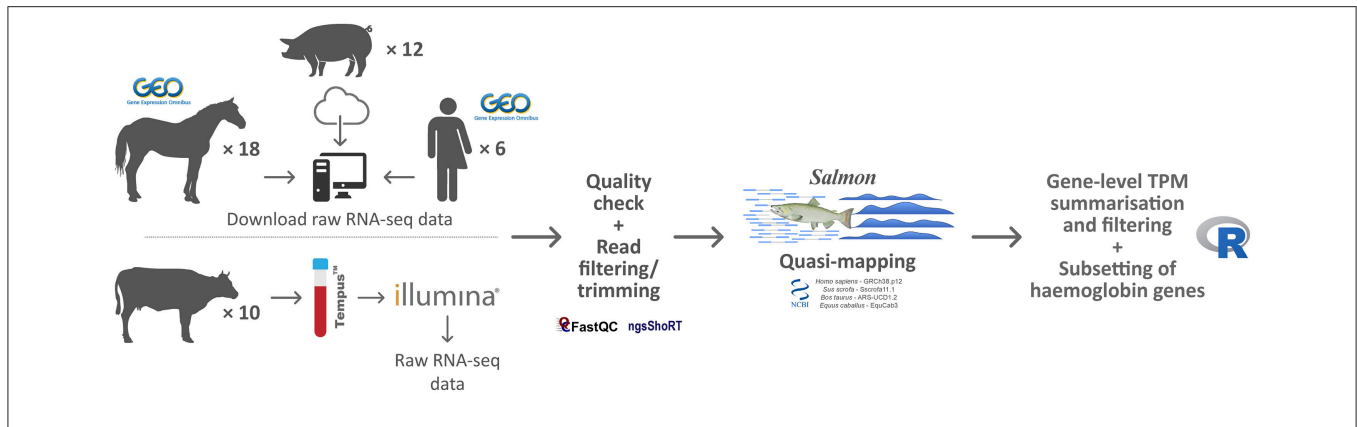


FIGURE 1 | Schematic of the bioinformatics workflow for RNA-seq data acquisition, quality control, analysis, and interpretation.

Porcine peripheral blood samples were collected from 12 healthy crossbred pigs [Duroc \times (Landrace \times Yorkshire)] using Tempus™ blood RNA tubes (Applied Biosystems™/Thermo Fisher Scientific, Warrington, UK) and total RNA was purified using the MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit (Invitrogen™/Thermo Fisher Scientific) (Choi et al., 2014). Porcine *HBA* and *HBB* mRNA transcripts were subsequently depleted from a subset of the total RNA samples using a modified RNase H globin depletion method with custom porcine-specific antisense oligonucleotides for *HBA* and *HBB*. RNA-seq data was then generated from 24 PE RNA-seq libraries (12 undepleted and 12 globin-depleted).

Equine peripheral blood samples were collected using Tempus™ blood RNA tubes from 12 healthy Arabian horses (five females and seven males) at three different time points during flat racing training (Ropka-Molik et al., 2017). In addition, peripheral blood samples were collected from six healthy untrained Arabian horses (two females and four males). Total RNA was purified using the MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit and 37 of the 42 total RNA samples were used to generate single-end (SE) libraries for RNA-seq data generation. Globin depletion for the equine samples was not performed prior to RNA-seq library preparation (Katarzyna Ropka-Molik, pers. comm.).

Bovine Peripheral Blood Collection and RNA Extraction

Approximately 3 ml of peripheral blood from 10 age-matched healthy male Holstein-Friesian calves were collected into Tempus™ blood RNA tubes. The Tempus™ Spin RNA Isolation Kit (Applied Biosystems™/Thermo Fisher Scientific) was used to perform total RNA extraction and purification, following the manufacturer's instructions. RNA quantity and quality checking were performed using a NanoDrop™ 1,000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and an Agilent 2,100 Bioanalyzer using an RNA 6,000 Nano LabChip kit (Agilent Technologies Ltd., Cork, Ireland). The majority of samples displayed a 260/280 ratio >1.8 and an RNA integrity number (RIN) >8.0 (**Supplementary Table 2**). Globin mRNA depletion

was not performed on the total RNA samples purified from bovine peripheral blood samples.

Bovine RNA-Seq Library Generation and Sequencing

Individually barcoded strand-specific RNA-seq libraries were prepared with 1 µg of total RNA from each sample. Two rounds of poly(A)⁺ RNA purification were performed for all RNA samples using the Dynabeads® mRNA DIRECT™ Micro Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The purified poly(A)⁺ RNA was then used to generate strand-specific RNA-seq libraries using the ScriptSeq™ v2 RNA-Seq Library Preparation Kit, the ScriptSeq™ Index PCR Primers (Sets 1 to 4) and the FailSafe™ PCR enzyme system (all sourced from Epicentre®/Illumina® Inc., Madison, WI, USA), according to the manufacturer's instructions.

RNA-seq libraries were purified using the Agencourt® AMPure® XP system (Beckman Coulter Genomics, Danvers, MA, USA) according to the manufacturer's instructions for double size selection (0.75× followed by 1.0× ratio). RNA-seq libraries were quantified using a Qubit® fluorometer and Qubit® dsDNA HS Assay Kit (Invitrogen™/Thermo Fisher Scientific), while library quality checks were performed using an Agilent 2,100 Bioanalyzer and High Sensitivity DNA Kit (Agilent Technologies Ltd.). Individually barcoded RNA-seq libraries were pooled in equimolar quantities and the quantity and quality of the final pooled libraries (three pools in total) were assessed as described above. Cluster generation and high-throughput sequencing of three pooled RNA-seq libraries were performed using an Illumina® HiSeq™ 2,000 Sequencing System at the MSU Research Technology Support Facility (RTSF) Genomics Core (<https://rtsf.natsci.msu.edu/genomics>; Michigan State University, MI, USA). Each of the three pooled libraries were sequenced independently on five lanes split across multiple Illumina® flow cells. The pooled libraries were sequenced as PE 2 × 100 nucleotide reads using Illumina® version 5.0 sequencing kits.

Deconvolution (filtering and segregation of sequence reads based on the unique RNA-seq library barcode index

sequences; **Supplementary Table 2**) was performed by the MSU RTSF Genomics Core using a pipeline that simultaneously demultiplexed and converted pooled sequence reads into discrete FASTQ files for each RNA-seq sample with no barcode index mismatches permitted. The RNA-seq FASTQ sequence read data for the bovine samples were obtained from the MSU RTSF Genomics Core FTP server.

RNA-Seq Data Quality Control and Filtering/Trimming of Reads

Bioinformatics procedures and analyses were performed as described below for the human, porcine, equine, and bovine samples, except where specifically indicated. All of the bioinformatics workflow scripts were developed using GNU bash (version 4.3.48) (Free Software Foundation, 2013), Python (version 3.5.2) (Python Software Foundation, 2017), and R (version 3.4.0) (R Core Team, 2017). The scripts and further information are available at a public GitHub repository (https://github.com/carolcorreia/Globin_RNA-sequencing).

Computational analyses were performed on a 32-core Linux Compute Server (4× AMD Opteron™ 6220 processors at 3.0 GHz with 8 cores each), with 256 GB of RAM, 24 TB of hard disk drive storage, and with Ubuntu Linux OS (version 14.04.4 LTS). Deconvoluted FASTQ files (generated from SE equine RNA-seq libraries and PE RNA-seq libraries for the other species) were quality-checked with FastQC (version 0.11.5) (Andrews, 2016).

Using the ngsShoRT software package (version 2.2) (Chen et al., 2014), filtering/trimming consisted of: (1) removal of SE or PE reads with adapter sequences (with up to three mismatches); (2) removal of SE or PE reads of poor quality (i.e., at least one of the reads containing $\geq 25\%$ bases with a Phred quality score below 20); (3) for porcine samples only, 10 bases were trimmed at the 3' end of all reads; (4) removal of SE or PE reads that did not meet the required minimum length (70 nucleotides for human and equine, 80 nucleotides for porcine and 100 nucleotides for bovine). Filtered/trimmed FASTQ files were then re-evaluated using FastQC. Filtered FASTQ files were transferred to a 36-core/64-thread Compute Server (2× Intel® Xeon® CPU E5-2697 v4 at 2.30 GHz with 18 cores each), with 512 GB of RAM, 96 TB SAS storage (12× 8 TB at 7200 rpm), 480 GB SSD storage, and with Ubuntu Linux OS (version 16.04.2 LTS).

Transcript Quantification

The Salmon software package (version 0.8.2) (Patro et al., 2017) was used in quasi-mapping-mode for transcript quantification. Sequence-specific and fragment-level GC bias correction was enabled and transcript abundance was quantified in transcripts per million (TPM) for each filtered library (multiple lanes from the same library were processed together) was estimated after mapping of SE or PE reads to their respective reference transcriptomes. As summarised in **Table 1**, the NCBI RefSeq database is currently the only one to contain haemoglobin gene annotations for all species analysed. Hence, NCBI RefSeq reference transcript models were used for the human, porcine, equine, and bovine data sets. Detailed information about these reference transcriptomes is provided in **Supplementary Table 3**.

Gene Annotations and Summarisation of TPM Estimates at the Gene Level

Using R (3.5.0) within the RStudio IDE (version 1.1.447) (R Studio Team, 2015) and Bioconductor (version 3.7 using BiocInstaller 1.30.0) (Gentleman et al., 2004), the GenomicFeatures (version 1.32.0) (Lawrence et al., 2013) and AnnotationDbi (version 1.42.1) (Pagès et al., 2017) packages were used to obtain corresponding gene and transcript identifiers from the NCBI RefSeq annotation releases pertinent to each species, as detailed in **Table 1**. Using these identifiers, the tximport (version 1.8.0) package (Soneson et al., 2015) was used to import into R and summarise at gene level the TPM estimates obtained from the Salmon tool. A threshold of greater than or equal to 1 TPM across at least half of the total number of samples (≥ 12 for human and porcine, ≥ 18 for equine, and ≥ 5 for bovine) was applied in order to remove lowly expressed genes.

Data Exploration, Plotting, and Summary Statistics

Data wrangling and tidying from all species was performed using the following R packages: tidyverse (version 1.2.1) (Wickham, 2017b), dplyr (version 0.7.5) (Wickham et al., 2017), tidyr (version 0.8.1) (Wickham and Henry, 2017), reshape2 (version 1.4.3) (Wickham, 2017a), and magrittr (version 1.5) (Bache and Wickham, 2017). The ggplot2 (version 2.2.1) (Wickham and Chang, 2017), and ggjoy (version 0.4.1) (Wilke, 2017), packages were used for figure generation. Finally, the mean and standard deviation were calculated for the undepleted and globin-depleted groups in each species using the skimr (version 1.0.2) R package (McNamara et al., 2017).

RESULTS AND DISCUSSION

Status of Human, Porcine, Equine, and Bovine Haemoglobin Gene Annotations

Annotation of the haemoglobin subunit alpha 1 and 2 genes (*HBA1* and *HBA2*, respectively) is well-established for the human genome; however, annotations for these genes in the porcine, equine, and bovine genomes are inconsistent across databases. As shown in **Table 1**, the porcine *HBA* gene annotation is absent from Ensembl and the UCSC Table Browser. For the NCBI RefSeq database, this gene has been assigned to two loci (*LOC110259958* and *LOC100737768*) that have similar descriptions (haemoglobin subunit alpha and haemoglobin subunit alpha-like). Therefore, these NCBI LOC symbols were used.

Equine *HBA* (*HBA1*) and *HBA2* genes are absent from the current Ensembl annotation release. Similarly, bovine *HBA1* and *HBA* (*HBA2*) have been annotated as *GLNC1* in Ensembl, whereas *HBA1* is absent from the UCSC Table Browser annotation (**Table 1**). In the NCBI RefSeq database, equine *HBA* (*HBA1*) is described as haemoglobin subunit alpha 1; and bovine *HBA* (*HBA2*) is described as haemoglobin subunit alpha 2, thus their descriptions are shown in parenthesis herein. In contrast to these observations, haemoglobin subunit beta (*HBB*) genes for

TABLE 1 | Status of current human, porcine, equine, and bovine haemoglobin gene annotations in the Ensembl, NCBI RefSeq, and UCSC databases.

| | Ensembl | NCBI RefSeq | UCSC table browser |
|---|--|---|--|
| <i>Homo sapiens</i> | | | |
| Annotation release | Human release 92 (April 2018) ^a | NCBI <i>Homo sapiens</i> Annotation Release 109 (March 2018) ^b | hg38.refGene annotation track (last updated on May 2018) ^{c,d,e} |
| Genome assembly used to derive annotation | GRCh38.p12, GCA_000001405.27, December 2017 | GRCh38.p12, GCF_000001405.38, December 2017 | GRCh38, GCF_000001405.15, December 2013 |
| <i>HBA1</i> | Annotated with gene ID ENSG00000206172 | Annotated with Entrez Gene ID 3039 | Annotated with Entrez Gene ID 3039 |
| <i>HBA2</i> | Annotated with gene ID ENSG00000188536 | Annotated with Entrez Gene ID 3040 | Annotated with Entrez Gene ID 3040 |
| <i>HBB</i> | Annotated with gene ID ENSG00000244734 | Annotated with Entrez Gene ID 3043 | Annotated with Entrez Gene ID 3043 |
| <i>Sus scrofa</i> | | | |
| Annotation release | Pig release 92 (April 2018) ^a | NCBI <i>Sus scrofa</i> Annotation Release 106 (May 2017) ^b | susScr3.refGene annotation track (last updated on May 2018) ^{d,e} |
| Genome assembly used to derive annotation | Sscrofa11.1, GCA_000003025.6, February 2017 | Sscrofa11.1, GCF_000003025.6, February 2017 | Sscrofa11.1, GCF_000003025.6, February 2017 |
| <i>LOC110259958 (HBA)</i> | Absent from current annotation release, accessible via online search with ID 110259958.1 | Annotated with Entrez Gene ID 110259958 | Absent |
| <i>LOC100737768 (HBA)</i> | Absent from current annotation release, accessible via online search with ID 100737768.1 | Annotated with Entrez Gene ID 100737768 | Absent |
| <i>HBB</i> | Annotated with gene ID ENSSCG00000014725 | Annotated with Entrez Gene ID 407066 | Annotated with Entrez Gene ID 407066 |
| <i>Equus caballus</i> | | | |
| Annotation release | Horse release 92 (April 2018) ^a | NCBI <i>Equus caballus</i> Annotation Release 103 (January 2018) ^b | equCab2.refGene annotation track (last updated on May 2018) ^{d,e} |
| Genome assembly used to derive annotation | EquCab 2, GCA_000002305.1, September 2007 | EquCab3, GCF_002863925.1, May 2018 | EquCab 2, GCA_000002305.1, September 2007 |
| <i>HBA</i> (also known as <i>HBA1</i>) | Absent | Annotated with Entrez Gene ID 100036557 | Annotated with Entrez Gene ID 100036557 |
| <i>HBA2</i> | Absent | Annotated with Entrez Gene ID 100036558 | Annotated with Entrez Gene ID 100036558 |
| <i>HBB</i> | Annotated with gene ID ENSECAG00000010020 | Annotated with Entrez Gene ID 100054109 | Annotated with Entrez Gene ID 100054109 |
| <i>Bos taurus</i> | | | |
| Annotation release | Cow release 92 (April 2018) ^a | NCBI <i>Bos taurus</i> Annotation Release 106 (May 2018) ^b | bosTau8.refGene annotation track (last updated on May 2018) ^{d,e} |
| Genome assembly used to derive annotation | UMD3.1, GCA_000003055.3, November 2009 | ARS-UCD1.2, GCF_002263795.1, April 2018 | UMD3.1.1, GCA_000003055.4, June 2014 |
| <i>HBA1</i> | Annotated as <i>GLNC1</i> with ID ENSBTAG00000026417 | Annotated with Entrez gene ID 100140149 | Absent |
| <i>HBA</i> (also known as <i>HBA2</i>) | Annotated as <i>GLNC1</i> with ID ENSBTAG00000026418 | Annotated with Entrez Gene ID 512439 | Annotated with Entrez Gene ID 512439 |
| <i>HBB</i> | Annotated with gene ID ENSBTAG00000038748 | Annotated with Entrez Gene ID 280813 | Annotated with Entrez Gene ID 280813 |

^a(Zerbino et al., 2018).^b(O'Leary et al., 2016).^c(Kent et al., 2002).^d(Karolchik et al., 2004).^e(Tyner et al., 2017).

the four species are well-annotated in Ensembl, NCBI RefSeq and UCSC Genome Browser databases (Table 1).

At the time of writing, NCBI RefSeq is the only database that contains annotations for all three haemoglobin genes in all species analysed. Additionally, equine and bovine gene annotations are based on the latest genome assemblies (Table 1). EquCab3 and ARS-UCD1.2 have incorporated major improvements compared to previous versions, including increased genome coverage (from 6.8× and 9×, to 80×,

respectively), and incorporation of PacBio sequencing reads (Kalbfleisch et al., 2018; Rosen et al., 2018).

Basic RNA-Seq Data Outputs

Unfiltered SE (equine libraries) or PE (human, porcine, and bovine libraries) RNA-seq FASTQ files were quality-checked, adapter- and quality-filtered prior to transcript quantification. As shown in Table 2, the human and porcine undepleted groups

each had ~40 million (M) raw reads per library, whereas globin-depleted libraries showed a mean of ~37 and 31 M, respectively. Equine and bovine libraries, which did not include a globin depletion step had an average of 24 M raw reads and 21 M raw read pairs, respectively.

After adapter- and quality-filtering of RNA-seq libraries, an average of 20 and 29% read pairs were removed from the human undepleted and globin-depleted libraries, respectively. Conversely, ~12% of read pairs were removed from each of the porcine undepleted and globin-depleted libraries. For the undepleted equine and bovine RNA-seq libraries, an average of 0.2% reads and 17% read pairs were removed, respectively. Detailed information on filtering/trimming of RNA-seq libraries from all species, including technical replicates from libraries sequenced over multiple lanes, is presented in **Supplementary Table 4**. All data sets exhibited a mean mapping rate >70% (**Table 2**). **Supplementary Tables 5** contain sample-specific RNA-seq mapping statistics.

Transcript Quantification

Transcript-level TPM estimates generated using the Salmon tool were imported into the R environment and summarised at gene level with the package tximport (Soneson et al., 2015). Gene-level TPM estimates represent the sum of corresponding transcript-level TPMs and provide results that are more accurate and comprehensible than transcript-level estimates (Soneson et al., 2015). In the current study, gene-level TPM estimates are referred as TPM.

Filtering of lowly expressed genes (see section Gene Annotations and Summarisation of TPM Estimates at the Gene Level) resulted in 12,951 genes expressed across all human samples, and represented 24% of 54,644 total annotated genes and pseudogenes. Porcine samples showed a total of 9,396 expressed genes (31% of 30,334 annotated genes and pseudogenes); and equine and bovine samples exhibited 12,724 (38% of 33,146) and 14,044 (40% of 35,143) expressed genes, respectively.

The density distribution of TPM values for the human and porcine samples improved after globin depletion; this is evident by the shift of gene detection levels toward greater log₁₀ TPM values for the globin-depleted samples in **Figure 2**. In this regard, it is noteworthy that the undepleted bovine and equine samples also exhibited similar TPM density distributions to the human and porcine globin-depleted samples.

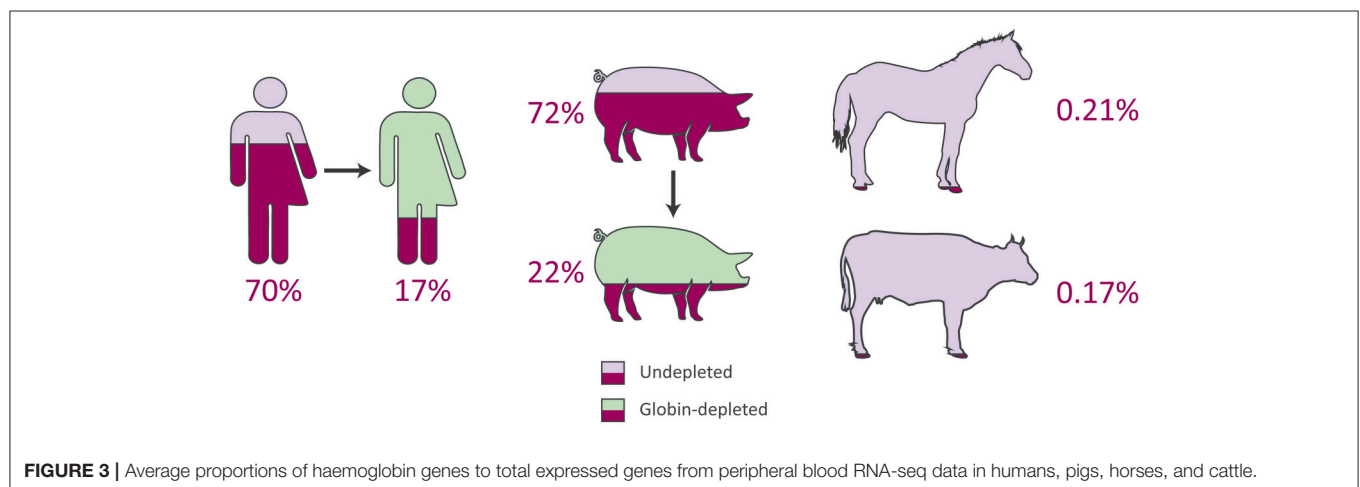
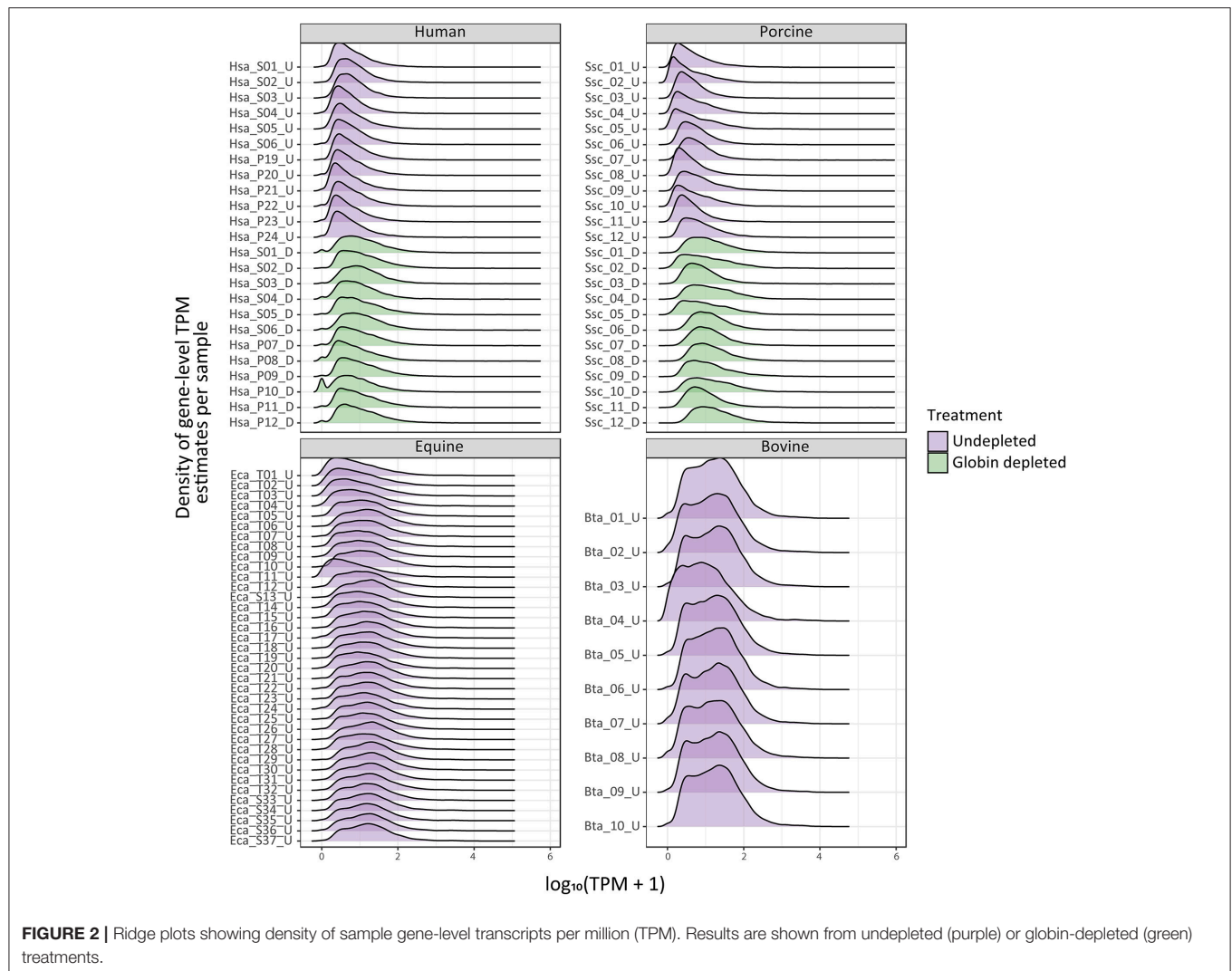
Proportions of Human and Porcine Haemoglobin Gene Transcripts in Undepleted and Depleted Peripheral Blood

In line with previous reports (Field et al., 2007; Mastrokolias et al., 2012), the proportion of haemoglobin gene transcripts (*HBA1*, *HBA2*, and *HBB*) detected in undepleted human peripheral blood samples for the current study averaged 70% (**Figure 3** and **Supplementary Table 6**), which is lower than the mean proportion of 81% reported by Shin et al. (2014). On the other hand, after depletion the human samples exhibited an identical reduction to a 17% proportion of globin sequence reads in both

TABLE 2 | Summary of RNA-seq filtering/trimming and mapping statistics.

| Species | Treatment | RNA-seq library type | Sequencing mode | Mean no. of reads (SE) or pairs (PE) | Mean no. of reads (SE) or pairs (PE) removed | Mean proportion of reads (SE) or pairs (PE) removed | Mean no. of observed fragments* | Mean no. of mapped fragments* | Average mapping rate | Reference source |
|-----------------------|-----------------|-------------------------|-----------------|--------------------------------------|--|---|---------------------------------|-------------------------------|----------------------|------------------|
| <i>Homo sapiens</i> | Undepleted | Inward unstranded | PE | 40,218,886 | 8,203,217 | 20.4% | 32,015,669 | 25,593,239 | 80.9% | NCBI RefSeq |
| <i>Homo sapiens</i> | Globin depleted | Inward unstranded | PE | 36,874,759 | 10,704,088 | 29.0% | 26,170,671 | 20,371,624 | 75.8% | NCBI RefSeq |
| <i>Sus scrofa</i> | Undepleted | Inward unstranded | PE | 39,036,515 | 4,613,991 | 11.8% | 28,685,437 | 25,129,140 | 87.7% | NCBI RefSeq |
| <i>Sus scrofa</i> | Globin depleted | Inward unstranded | PE | 31,339,886 | 3,899,427 | 12.4% | 22,867,049 | 19,959,995 | 87.3% | NCBI RefSeq |
| <i>Equus caballus</i> | Undepleted | Unstranded | SE | 24,271,141 | 38,892 | 0.2% | 14,850,797 | 11,387,774 | 76.5% | NCBI RefSeq |
| <i>Bos taurus</i> | Undepleted | Inward stranded forward | PE | 20,495,983 | 3,474,597 | 17.0% | 17,021,386 | 12,353,147 | 72.6% | NCBI RefSeq |

*The Salmon tool categorises fragments as single read (for SE RNA-seq libraries) or a read pair (for PE RNA-seq libraries).



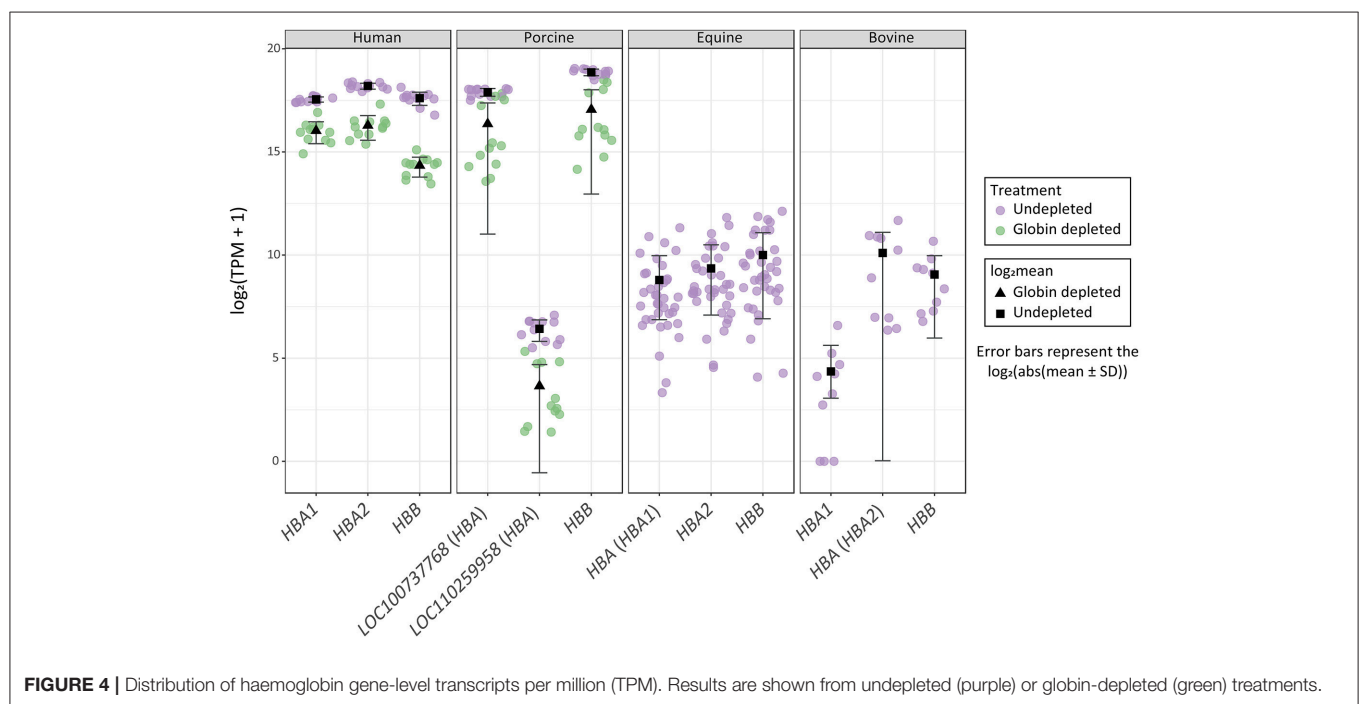
the present study and that of Shin et al. (2014) (**Figure 3** and **Supplementary Table 6**).

In the current study, for the undepleted porcine peripheral blood samples, the percentage of haemoglobin gene transcripts

(*LOC110259958* [*HBA*], *LOC100737768* [*HBA*], and *HBB*) observed as a proportion of the total expressed genes was 72% (**Figure 3** and **Supplementary Table 6**), which is considerably larger than the mean of 46.1% reported in the original study

TABLE 3 | Summary statistics for haemoglobin gene-level transcripts per million (TPM).

| Species | Gene symbol | Treatment | No. of samples | Mean TPM | Standard deviation |
|-----------------------|---------------------------|-----------------|----------------|----------|--------------------|
| <i>Homo sapiens</i> | <i>HBA1</i> | Undepleted | 12 | 191,209 | 16,601 |
| <i>Homo sapiens</i> | <i>HBA1</i> | Globin depleted | 12 | 66,718 | 23,557 |
| <i>Homo sapiens</i> | <i>HBA2</i> | Undepleted | 12 | 300,000 | 29,523 |
| <i>Homo sapiens</i> | <i>HBA2</i> | Globin depleted | 12 | 79,818 | 31,259 |
| <i>Homo sapiens</i> | <i>HBB</i> | Undepleted | 12 | 200,000 | 43,262 |
| <i>Homo sapiens</i> | <i>HBB</i> | Globin depleted | 12 | 20,770 | 6,706 |
| <i>Sus scrofa</i> | <i>LOC110259958 (HBA)</i> | Undepleted | 12 | 86 | 30 |
| <i>Sus scrofa</i> | <i>LOC110259958 (HBA)</i> | Globin depleted | 12 | 13 | 13 |
| <i>Sus scrofa</i> | <i>LOC100737768 (HBA)</i> | Undepleted | 12 | 243,864 | 31,605 |
| <i>Sus scrofa</i> | <i>LOC100737768 (HBA)</i> | Globin depleted | 12 | 84,021 | 86,095 |
| <i>Sus scrofa</i> | <i>HBB</i> | Undepleted | 12 | 476,284 | 52,939 |
| <i>Sus scrofa</i> | <i>HBB</i> | Globin depleted | 12 | 136,172 | 128,232 |
| <i>Equus caballus</i> | <i>HBA (HBA1)</i> | Undepleted | 37 | 443 | 560 |
| <i>Equus caballus</i> | <i>HBA2</i> | Undepleted | 37 | 653 | 789 |
| <i>Equus caballus</i> | <i>HBB</i> | Undepleted | 37 | 1,024 | 1,144 |
| <i>Bos taurus</i> | <i>HBA1</i> | Undepleted | 10 | 21 | 29 |
| <i>Bos taurus</i> | <i>HBA (HBA2)</i> | Undepleted | 10 | 1,101 | 1,102 |
| <i>Bos taurus</i> | <i>HBB</i> | Undepleted | 10 | 532 | 469 |



(Choi et al., 2014). Similarly, after depletion, the porcine samples in the present study contained a mean proportion of 22% globin transcripts (Figure 3 and Supplementary Table 6) compared to a mean proportion of 8.9% reported by Choi et al. (2014). Additionally, Table 3 shows the mean TPM for each haemoglobin gene across undepleted or globin-depleted samples.

A number of possible explanations, including the different approaches used for read mapping and transcript quantification, may account for the different proportions of haemoglobin gene

transcript detected in human and porcine samples for the present study compared to the original studies (Choi et al., 2014; Shin et al., 2014). For the present study, a recently developed lightweight alignment method was adopted (Salmon and tximport), in contrast to the more traditional methodologies used in the original publications. Shin et al. (2014) used the TopHat and Cufflinks software tools (Trapnell et al., 2012), while Choi et al. (2014) implemented TopHat with Htseq-count (Anders et al., 2015). In addition to this, different gene

annotations were used: NCBI *Homo sapiens* Annotation Release 109 and NCBI *Sus scrofa* Annotation Release 106 were used for the present study, while UCSC hg18 (*H. sapiens*) and Ensembl release 71 (*S. scrofa*) were used by Shin et al. (2014) and Choi et al. (2014), respectively.

Equine and Bovine Peripheral Blood Contains Extremely Low Levels of Haemoglobin Gene Transcripts

The equine and bovine peripheral blood samples, which did not undergo globin depletion, had extremely low proportions of haemoglobin gene transcripts to total expressed genes: 0.21 and 0.17%, respectively (Figure 3 and Supplementary Table 6). Notably, similar results have been reported in a transcriptomics study of bovine peripheral blood in response to vaccination against neonatal pancytopenia. In that study, 12 cows were profiled before and after vaccination (24 peripheral blood samples in total), and a mean proportion of 1.0% of RNA-seq reads were observed to map to the bovine α haemoglobin gene cluster on BTA25 or to the β haemoglobin gene cluster on BTA15 (Demasius et al., 2013). To the best of our knowledge, this is the first time that the average number of equine haemoglobin transcripts have been reported for RNA-seq data.

Finally, it is important to note that log₂ TPM values for haemoglobin gene transcripts in the undepleted equine and bovine peripheral blood RNA samples are substantially lower than log₂ TPM values for the globin-depleted human and porcine peripheral blood RNA samples (Figure 4). This is a direct consequence of extremely low levels of circulating reticulocytes in equine and bovine peripheral blood (Tablin and Weiss, 1985; Harper et al., 1994; Hossain et al., 2003; Cooper et al., 2005).

CONCLUSION

In light of our RNA-seq data analyses, we propose that globin mRNA transcript depletion is not a pre-requisite for transcriptome profiling of bovine and equine peripheral blood samples. This observation greatly simplifies the laboratory and bioinformatics workflows required for RNA-seq studies of whole blood collected from domestic cattle and horses. It will also be directly relevant to future work on blood-based biomarker and biosignature development in the context of infectious disease, reproduction, nutrition, and animal welfare. For example, transcriptomics of peripheral blood has been used extensively in development of new diagnostic and prognostic modalities for human tuberculosis (HTB) disease caused by infection with *Mycobacterium tuberculosis* (for reviews see: Blankley et al., 2014; Haas et al., 2016; Weiner and Kaufmann, 2017; Goletti et al., 2018). Therefore, as a consequence of this HTB research, comparable transcriptomics studies in cattle (Meade et al., 2007; Killick et al., 2011; Blanco et al., 2012; Churbanov and Milligan, 2012; McLoughlin et al., 2014; Cheng et al., 2015), and the ease with which RNA-seq can be performed in bovine peripheral blood, it should be feasible to develop transcriptomics-based biomarkers and biosignatures for bovine tuberculosis caused by *M. bovis* infection.

DATA ACCESSIBILITY

The RNA-seq data generated for this study using peripheral blood from 10 age-matched healthy male Holstein-Friesian calves can be obtained from the ENA database (PRJEB27764).

ETHICS STATEMENT

Animal experimental work for the present study (cattle samples) was carried out according to the UK Animal (Scientific Procedures) Act 1986. The study protocol was approved by the Animal Health and Veterinary Laboratories Agency (AHVLA–Weybridge, UK), now the Animal & Plant Health Agency (APHA), Animal Use Ethics Committee (UK Home Office PCD number 70/6905).

AUTHOR CONTRIBUTIONS

DEM, SG, CC, and KM conceived and designed the project and organised bovine sample collection. KM, NN, DAM, and JB performed RNA extraction and RNA-seq library generation. CC, KM, NN, KR-A, and DEM performed the analyses. CC and DEM wrote the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by Investigator Grants from Science Foundation Ireland (Nos: SFI/08/IN.1/B2038 and SFI/15/IA/3154), a Research Stimulus Grant from the Department of Agriculture, Food and the Marine (No: RSF 06 405), a European Union Framework 7 Project Grant (No: KBBE-211602- MACROSYS), a Brazilian Science Without Borders—CAPES grant (No: BEX-13070-13-4) and the UCD Wellcome Trust funded Computational Infection Biology Ph.D. Programme (Grant no: 097429/Z/11/Z).

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Prof Martin Vordermeier and Dr Bernardo Villarreal-Ramos (Animal and Plant Health Agency, UK) for provision of bovine peripheral blood samples, Prof Graham Plastow (University of Alberta, Canada) for provision of porcine peripheral blood RNA-seq data. We also thank Drs. Gabriella Farries (University College Dublin) and Kerri Malone (EMBL-EBI, Cambridge, UK) for stimulating discussion and advice concerning genome annotations, equine genetics and data visualisation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00278/full#supplementary-material>

REFERENCES

- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Andrews, S. (2016). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Animal Task Force (2016). *A Strategic Research and Innovation Agenda for a Sustainable livestock sector in Europe*. Available online at: <http://www.animaltaskforce.eu>
- Bache, S. M., and Wickham, H. (2017). *magrittr: A Forward-Pipe Operator for R*. Available online at: <https://github.com/tidyverse/magrittr>
- Blanco, F. C., Soria, M., Bianco, M. V., and Bigi, F. (2012). Transcriptional response of peripheral blood mononuclear cells from cattle infected with *Mycobacterium bovis*. *PLoS ONE* 7:e41066. doi: 10.1371/journal.pone.0041066
- Blankley, S., Berry, M. P., Graham, C. M., Bloom, C. I., Lipman, M., and O'Garra, A. (2014). The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 369:20130427. doi: 10.1098/rstb.2013.0427
- Bowyer, J. F., Tranter, K. M., Hanig, J. P., Crabtree, N. M., Schleimer, R. P., and George, N. I. (2015). Evaluating the stability of RNA-seq transcriptome profiles and drug-induced immune-related expression changes in whole blood. *PLoS ONE* 10:e0133315. doi: 10.1371/journal.pone.0133315
- Bruder, C. E., Yao, S., Larson, F., Camp, J. V., Tapp, R., McBrayer, A., et al. (2010). Transcriptome sequencing and development of an expression microarray platform for the domestic ferret. *BMC Genomics* 11:251. doi: 10.1186/1471-2164-11-251
- Chaussabel, D. (2015). Assessment of immune status using blood transcriptomics and potential implications for global health. *Semin. Immunol.* 27, 58–66. doi: 10.1016/j.smim.2015.03.002
- Chen, C., Khaleel, S. S., Huang, H., and Wu, C. H. (2014). Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol. Med.* 9:8. doi: 10.1186/1751-0473-9-8
- Cheng, Y., Chou, C.-H., and Tsai, H.-J. (2015). *In vitro* gene expression profile of bovine peripheral blood mononuclear cells in early *Mycobacterium bovis* infection. *Exp. Ther. Med.* 10, 2102–2118. doi: 10.3892/etm.2015.2814
- Choi, I., Bao, H., Kommadath, A., Hosseini, A., Sun, X., Meng, Y., et al. (2014). Increasing gene discovery and coverage using RNA-seq of globin RNA reduced porcine blood samples. *BMC Genomics* 15:954. doi: 10.1186/1471-2164-15-954
- Churbanov, A., and Milligan, B. (2012). Accurate diagnostics for bovine tuberculosis based on high-throughput sequencing. *PLoS ONE* 7:e50147. doi: 10.1371/journal.pone.0050147
- Cooper, C., Sears, W., and Bienzle, D. (2005). Reticulocyte changes after experimental anemia and erythropoietin treatment of horses. *J. Appl. Physiol.* 99, 915–921. doi: 10.1152/japplphysiol.00438.2005
- de Greeff, A., Bikker, P., Smit-Heinsbroek, A., Bruininx, E., Zwolschen, H., Fijten, H., et al. (2016). Increased fat and polyunsaturated fatty acid content in sow gestation diet has no effect on gene expression in progeny during the first 7 days of life. *J. Anim. Physiol. Anim. Nutr.* 100, 127–135. doi: 10.1111/jpn.12345
- Debey, S., Schoenbeck, U., Hellmich, M., Gathof, B. S., Pillai, R., Zander, T., et al. (2004). Comparison of different isolation techniques prior gene expression profiling of blood derived cells: impact on physiological responses, on overall expression and the role of different cell types. *Pharmacogenomics J.* 4, 193–207. doi: 10.1038/sj.tpj.6500240
- Demasius, W., Weikard, R., Hadlich, F., Muller, K. E., and Kuhn, C. (2013). Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle. *Vet. Res.* 44:93. doi: 10.1186/1297-9716-44-93
- Elgendy, R., Giantin, M., Castellani, F., Grotta, L., Palazzo, F., Dacasto, M., et al. (2016). Transcriptomic signature of high dietary organic selenium supplementation in sheep: a nutrigenomic insight using a custom microarray platform and gene set enrichment analysis. *J. Anim. Sci.* 94, 3169–3184. doi: 10.2527/jas.2016-0363
- Fan, H., and Hegde, P. S. (2005). The transcriptome in blood: challenges and solutions for robust expression profiling. *Curr. Mol. Med.* 5, 3–10. doi: 10.2174/1566524053152861
- Field, L. A., Jordan, R. M., Hadix, J. A., Dunn, M. A., Shriver, C. D., Ellsworth, R. E., et al. (2007). Functional identity of genes detectable in expression profiling assays following globin mRNA reduction of peripheral blood samples. *Clin. Biochem.* 40, 499–502. doi: 10.1016/j.clinbiochem.2007.01.004
- Free Software Foundation (2013) *Bash (4.3.48) Unix Shell Program*. Available online at: <http://ftp.gnu.org/gnu/bash/>
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80. doi: 10.1186/gb-2004-5-10-r80
- Goletti, D., Lee, M. R., Wang, J. Y., Walter, N., and Ottenhoff, T. H. M. (2018). Update on tuberculosis biomarkers: from correlates of risk, to correlates of active disease and of cure from disease. *Respirology* 23, 455–466. doi: 10.1111/resp.13272
- Greer, J. P., Arber, D. A., Glader, B., List, A. F., Means, R. T., Paraskevas, F., et al. (2013). *Wintrobe's Clinical Hematology*. Philadelphia, PA: Lippincott, Williams and Wilkins
- Haas, C. T., Roe, J. K., Pollara, G., Mehta, M., and Noursadeghi, M. (2016). Diagnostic 'omics' for active tuberculosis. *BMC Med.* 14:37. doi: 10.1186/s12916-016-0583-9
- Harper, S. B., Hurst, W. J., Ohlsson-Wilhelm, B., and Lang, C. M. (1994). The response of various hematologic parameters in the young bovine subjected to multiple phlebotomies. *ASAIO J.* 40, M816–M825.
- Holcomb, Z. E., Tsalik, E. L., Woods, C. W., and McClain, M. T. (2017). Host-based peripheral blood gene expression analysis for diagnosis of infectious diseases. *J. Clin. Microbiol.* 55, 360–368. doi: 10.1128/jcm.01057-16
- Hossain, M. A., Yamato, O., Yamasaki, M., Otsuka, Y., and Maeda, Y. (2003). Relation between reticulocyte count and characteristics of erythrocyte 5'-nucleotidase in dogs, cats, cattle and humans. *J. Vet. Med. Sci.* 65, 193–197. doi: 10.1292/jvms.65.193
- Huang, Z., Gallot, A., Lao, N. T., Puechmaile, S. J., Foley, N. M., Jebb, D., et al. (2016). A nonlethal sampling method to obtain, generate and assemble whole blood transcriptomes from small, wild mammals. *Mol. Ecol. Resour.* 16, 150–162. doi: 10.1111/1755-0998.12447
- Jégou, M., Gondret, F., Vincent, A., Trefeu, C., Gilbert, H., and Louveau, I. (2016). Whole blood transcriptomics is relevant to identify molecular changes in response to genetic selection for feed efficiency and nutritional status in the pig. *PLoS ONE* 11:e0146550. doi: 10.1371/journal.pone.0146550
- Kalbfleisch, T. S., Rice, E., DePriest, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., et al. (2018). EquCab3, an updated reference genome for the domestic horse. *bioRxiv[Preprint]*. 306928. doi: 10.1101/306928
- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., et al. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496. doi: 10.1093/nar/gkh103
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006. doi: 10.1101/gr.229102
- Killick, K. E., Browne, J. A., Park, S. D., Magee, D. A., Martin, I., Meade, K. G., et al. (2011). Genome-wide transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis* reveals suppression of host immune genes. *BMC Genomics* 12:611. doi: 10.1186/1471-2164-12-611
- Ko, E. R., Yang, W. E., McClain, M. T., Woods, C. W., Ginsburg, G. S., and Tsalik, E. L. (2015). What was old is new again: using the host response to diagnose infectious disease. *Expert Rev. Mol. Diagn.* 15, 1143–1158. doi: 10.1586/14737159.2015.1059278
- Kolli, V., Upadhyay, R. C., and Singh, D. (2014). Peripheral blood leukocytes transcriptomic signature highlights the altered metabolic pathways by heat stress in zebu cattle. *Res. Vet. Sci.* 96, 102–110. doi: 10.1016/j.rvsc.2013.11.019
- Krjutskov, K., Koel, M., Roost, A. M., Katayama, S., Einarsdottir, E., Jouhilahti, E. M., et al. (2016). Globin mRNA reduction for whole-blood transcriptome sequencing. *Sci. Rep.* 6:31584. doi: 10.1038/srep31584
- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118
- Liu, J., Walter, E., Stenger, D., and Thach, D. (2006). Effects of globin mRNA reduction methods on gene expression profiles from whole blood. *J. Mol. Diagn.* 8, 551–558. doi: 10.2353/jmoldx.2006.060021

- Mastrokolias, A., den Dunnen, J. T., van Ommen, G. B., t Hoen, P. A., and van Roon-Mom, W. M. (2012). Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics* 13:28. doi: 10.1186/1471-2164-13-28
- McLoughlin, K. E., Nalpas, N. C., Rue-Albrecht, K., Browne, J. A., Magee, D. A., Killick, K. E., et al. (2014). RNA-seq transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis*. *Front. Immunol.* 5:396. doi: 10.3389/fimmu.2014.00396
- McNamara, A., Arino de la Rubia, E., Zhu, H., Lowndes, J., Ellis, S., Waring, E., et al. (2017). *skimr: Compact and Flexible Summaries of Data*. Available online at: <https://github.com/ropenscilabs/skimr>
- Meade, K. G., Gormley, E., Doyle, M. B., Fitzsimons, T., O'Farrelly, C., Costello, E., et al. (2007). Innate gene repression associated with *Mycobacterium bovis* infection in cattle: toward a gene signature of disease. *BMC Genomics* 8:400. doi: 10.1186/1471-2164-8-400
- Mejias, A., and Ramilo, O. (2014). Transcriptional profiling in infectious diseases: ready for prime time? *J. Infect.* 68 (Suppl. 1), S94–99. doi: 10.1016/j.jinf.2013.09.018
- Morey, J. S., Neely, M. G., Lunardi, D., Anderson, P. E., Schwacke, L. H., Campbell, M., et al. (2016). RNA-seq analysis of seasonal and individual variation in blood transcriptomes of healthy managed bottlenose dolphins. *BMC Genomics* 17:720. doi: 10.1186/s12864-016-3020-8
- Nabarro, D., and Wannous, C. (2014). The potential contribution of livestock to food and nutrition security: the application of the One Health approach in livestock policy and practice. *Rev. Off. Int. Epizoot.* 33, 475–485. doi: 10.20506/rst.33.2.2292
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- O'Loughlin, A., Lynn, D. J., McGee, M., Doyle, S., McCabe, M., and Earley, B. (2012). Transcriptomic analysis of the stress response to weaning at housing in bovine leukocytes using RNA-seq technology. *BMC Genomics* 13:250. doi: 10.1186/1471-2164-13-250
- Pagès, H., Carlson, M., Falcon, S., and Li, N. (2017). *AnnotationDbi: Annotation Database Interface*. R package version 1.38.0. Available online at: <https://doi.org/10.18129/B9.bioc.AnnotationDbi>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Python Software Foundation (2017). *Python (version 3.5.2)*. Available online at: <https://www.python.org/downloads/release/python-352/>
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Available online at: <https://www.R-project.org>
- R Studio Team (2015). *RStudio: Integrated Development for R*. Available online at: <http://www.rstudio.com>
- Ramilo, O., and Mejias, A. (2009). Shifting the paradigm: host gene signatures for diagnosis of infectious diseases. *Cell Host Microbe* 6, 199–200. doi: 10.1016/j.chom.2009.08.007
- Ropka-Molik, K., Stefaniuk-Szmukier, M., Zukowski, K., Piorkowska, K., Gurgul, A., and Bugno-Poniewierska, M. (2017). Transcriptome profiling of Arabian horse blood during training regimens. *BMC Genet.* 18:31. doi: 10.1186/s12863-017-0499-1
- Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Zimin, A., et al. (2018). "Modernizing the bovine reference genome assembly," in *Proceedings of the World Congress on Genetics Applied to Livestock Production* (Auckland), 802.
- Schwochow, D., Serieys, L. E., Wayne, R. K., and Thalmann, O. (2012). Efficient recovery of whole blood RNA—a comparison of commercial RNA extraction protocols for high-throughput applications in wildlife species. *BMC Biotechnol.* 12:33. doi: 10.1186/1472-6750-12-33
- Shen, J., Zhou, C., Zhu, S., Shi, W., Hu, M., Fu, X., et al. (2014). Comparative transcriptome analysis reveals early pregnancy-specific genes expressed in peripheral blood of pregnant sows. *PLoS ONE* 9:e114036. doi: 10.1371/journal.pone.0114036
- Shin, H., Shannon, C. P., Fishbane, N., Ruan, J., Zhou, M., Balshaw, R., et al. (2014). Variation in RNA-seq transcriptome profiles of peripheral whole blood from healthy individuals with and without globin depletion. *PLoS ONE* 9:e91041. doi: 10.1371/journal.pone.0091041
- Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4:1521. doi: 10.12688/f1000research.7563.2
- Song, K. D., Dowd, S. E., Lee, H. K., and Kim, S. W. (2013). Long-term dietary supplementation of organic selenium modulates gene expression profiles in leukocytes of adult pigs. *Anim. Sci. J.* 84, 238–246. doi: 10.1111/j.1740-0929.2012.01060.x
- Tablin, F., and Weiss, L. (1985). Equine bone marrow: a quantitative analysis of erythroid maturation. *Anat. Rec.* 213, 202–206. doi: 10.1002/ar.1092130212
- Takahashi, J., Waki, S., Matsumoto, R., Odake, J., Miyaji, T., Tottori, J., et al. (2012). Oligonucleotide microarray analysis of dietary-induced hyperlipidemia gene expression profiles in miniature pigs. *PLoS ONE* 7:e37581. doi: 10.1371/journal.pone.0037581
- Thornton, P. K. (2010). Livestock production: recent trends, future prospects. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 2853–2867. doi: 10.1098/rstb.2010.0134
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Tyner, C., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Eisenhart, C., et al. (2017). The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res.* 45, D626–D634. doi: 10.1093/nar/gkw1134
- Weiner, J., and Kaufmann, S. H. (2017). High-throughput and computational approaches for diagnostic and prognostic host tuberculosis biomarkers. *Int. J. Infect. Dis.* 56, 258–262. doi: 10.1016/j.ijid.2016.10.017
- Wickham, H. (2017a). *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. Available online at: <https://github.com/hadley/reshape>
- Wickham, H. (2017b). *tidyverse*. Available online at: <http://tidyverse.tidyverse.org>
- Wickham, H., and Chang, W. (2017). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. Available online at: <http://ggplot2.tidyverse.org>
- Wickham, H., Francois, R., Henry, L., and Müller, K. (2017). *dplyr: A Grammar of Data Manipulation*. Available online at: <http://dplyr.tidyverse.org>
- Wickham, H., and Henry, L. (2017). *tidyr: Easily Tidy Data With 'spread()' and 'gather()' Functions*. Available online at: <http://tidyr.tidyverse.org>
- Wilke, C. O. (2017). *ggjoy: Joyplots in 'ggplot2'*. Available online at: <https://github.com/clauswilke/ggjoy>
- Winn, M. E., Zapala, M. A., Hovatta, I., Risbrough, V. B., Lillie, E., and Schork, N. J. (2010). The effects of globin on microarray-based gene expression analysis of mouse blood. *Mamm. Genome* 21, 268–275. doi: 10.1007/s00335-010-9261-y
- Wu, K., Miyada, G., Martin, J., and Finkelstein, D. (2003). *Technical Note: Globin Reduction Protocol: A Method for Processing Whole Blood RNA Samples for Improved Array Results*. Available online at: https://tools.thermofisher.com/content/sfs/brochures/blood2_technote.pdf
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhair, J., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* 46, D754–D761. doi: 10.1093/nar/gkx1098

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Correia, McLoughlin, Nalpas, Magee, Browne, Rue-Albrecht, Gordon and MacHugh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Analysis of a Commercial Egg Laying Line Challenged With Newcastle Disease Virus

Kaylee Rowland¹, Anna Wolc^{1,2}, Rodrigo A. Gallardo³, Terra Kelly^{3,4}, Huaijun Zhou⁴, Jack C. M. Dekkers¹ and Susan J. Lamont^{1*}

¹ Department of Animal Science, Iowa State University, Ames, IA, United States, ² Hy-Line International, Dallas Center, IA, United States, ³ School of Veterinary Medicine, University of California, Davis, Davis, CA, United States, ⁴ Department of Animal Science, University of California, Davis, Davis, CA, United States

OPEN ACCESS

Edited by:

Andrea B. Doeschl-Wilson,
University of Edinburgh,
United Kingdom

Reviewed by:

Yniv Palti,
Cool and Cold Water Aquaculture
Research (USDA-ARS), United States
Filippo Biscarini,
Consiglio Nazionale delle
Ricerche (CNR), Italy

*Correspondence:

Susan J. Lamont
sjlamont@iastate.edu

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 19 April 2018

Accepted: 30 July 2018

Published: 20 August 2018

Citation:

Rowland K, Wolc A, Gallardo RA,
Kelly T, Zhou H, Dekkers JCM and
Lamont SJ (2018) Genetic Analysis
of a Commercial Egg Laying Line
Challenged With Newcastle Disease
Virus. *Front. Genet.* 9:326.
doi: 10.3389/fgene.2018.00326

In low income countries, chickens play a vital role in daily life. They provide a critical source of protein through egg production and meat. Newcastle disease, caused by avian paramyxovirus type 1, has been ranked as the most devastating disease for scavenging chickens in Africa and Asia. High mortality among flocks infected with velogenic strains leads to a devastating loss of dietary protein and buying power for rural households. Improving the genetic resistance of chickens to Newcastle Disease virus (NDV), in addition to vaccination, is a practical target for improvement of poultry production in low income countries. Because response to NDV has a component of genetic control, it can be influenced through selective breeding. Adding genomic information to a breeding program can increase the amount of genetic progress per generation. In this study, we challenged a commercial egg-laying line with a lentogenic strain of NDV, measured phenotypic responses, collected genotypes, and associated genotypes with phenotypes. Collected phenotypes included viral load at 2 and 6 days post-infection (dpi), antibody levels pre-challenge and 10 dpi, and growth rates pre- and post-challenge. Six suggestive QTL associated with response to NDV and/or growth were identified, including novel and known QTL confirming previously reported associations with related traits. Additionally, previous RNA-seq analysis provided support for several of the genes located in or near the identified QTL. Considering the trend of negative genetic correlation between antibody and Newcastle Disease tolerance (growth under disease) and estimates of moderate to high heritability, we provide evidence that these NDV response traits can be influenced through selective breeding. Producing chickens that perform favorably in challenging environments will ultimately increase the supply of quality protein for human consumption.

Keywords: Newcastle disease virus, GWAS, poultry, disease challenge, genetic parameters, QTL, immune response

Abbreviations: dpi, days post-infection; EID₅₀, 50% embryo infectious dose; ELISA, enzyme-linked immunosorbent assay; GRM, genomic relationship matrix; Mb, mega base; NDV, Newcastle disease virus; QTL, quantitative trait loci; SE, standard error; SNP, single nucleotide polymorphism.

INTRODUCTION

In low income countries, chickens play a vital role in daily life. They provide important sources of high quality protein and macro and micronutrients. They are also important for livelihood and gender empowerment, as women are often the beneficiaries of poultry production, which is often not true with larger livestock (Guèye, 2000).

Newcastle disease, caused by avian paramyxovirus type 1, has been ranked as the most devastating disease for scavenging chickens (village chickens, allowed to roam with no to minimal feed provided) in Africa and Asia (Kitalyi, 1998). The more virulent strains of the virus can cause 80% mortality (number of deaths in the flock per infection event) in scavenging flocks (Kitalyi, 1998). High mortality among flocks lead to a devastating loss of dietary protein and buying power for rural households. Prevention of this disease through vaccination is challenging in rural, scavenging production systems. Difficulties arise in ensuring cold chain during transport of vaccines, inadequate vaccination programs, and high costs of administering booster vaccinations (Mayers et al., 2017). Improving the genetic resistance of chickens to NDV is a practical target for improvement of poultry production in low income countries.

Selective breeding has a demonstrated history of success in poultry production (Havenstein et al., 1994) and can be used to modulate many traits of chickens. Several reports have demonstrated genetic differences in response to NDV (Cole and Hutt, 1961; Gordon et al., 1970; Peleg et al., 1976; Soller et al., 1981; Pitcovski et al., 1987). Because response to NDV has a component of genetic control, it can be influenced through selective breeding. Adding genomic information to a breeding program can increase the amount of genetic progress per generation (Fulton, 2012). Recently, there has been a gap in the knowledge accumulation and study of NDV. However, the threat of NDV continues, as demonstrated by the 2018 outbreaks of virulent NDV in California. In this study, we challenged a commercial egg-laying line with a lentogenic (lowly virulent) strain of NDV, measured phenotypic responses, collected genotypes, and associated genotypes with phenotypes. We identified genomic regions associated with response to NDV and/or growth. A selective breeding program can be implemented, e.g., utilizing genomic information identified in this study, to produce chickens that perform favorably in challenging environments and ultimately increase the supply of quality protein for human consumption.

MATERIALS AND METHODS

Animals and Husbandry

The Iowa State University Institutional Animal Care and Use Committee approved all animal procedures and care in this study (log #1-13-7490-G). Pooled semen from 16 sires was used to inseminate 145 dams to produce 3 hatches of 200 mixed-sex chicks ($N = 600$) of a commercial brown egg laying line (Hy-line Brown, Hy-Line International). Birds were

provided *ad libitum* access to feed and water throughout the study period. Initially, 23 h of light was provided, which was gradually decreased to 13.5 h of light by day 29. Temperature at chick level on day of placement was 35°C and gradually decreased to 24°C by day 29 and held until completion of the experiment.

Experimental Design

On day of hatch, chicks were transported to a biosafety level II facility at Iowa State University. For each hatch, chicks were placed into one of three rooms, using pedigree information to distribute half-sibs into different rooms. At 21 days of age (0 dpi), birds were inoculated with 10^8 of 50% embryonic infectious dose (EID₅₀) of live attenuated type B1 LaSota strain NDV in a volume of 200 μ L. Virus propagation was detailed previously by Deist et al. (2017b). Virus was administered via a natural, ocular-nasal route. Each eye and nares received 50 μ L. Lachrymal fluid samples were collected to quantify viral load at 20, 23, and 27 days of age, hereafter designated as pre-challenge, 2 dpi and 6 dpi, respectively. Blood samples were collected to measure anti-NDV antibody levels on days 20 and 31, hereafter referred to as pre-challenge and 10 dpi, respectively. Body weights were recorded on days 0, 13, 21, 27, and 31 of age. The experimental design was performed across three replicates (3 hatches from the same dams and sires). In each replicate, 180 birds were challenged, 540 in total. The objective of this study is to find genotypic associations with quantitative responses to a viral challenge. Thus, pre-challenge measurements with confirmed null viral load serve as an internal control group.

Viral Load

To quantify viral load, viral RNA was isolated from lachrymal fluid and quantified via qPCR at three time points: pre-challenge ($n = 89$), 2 dpi ($n = 468$), and 6 dpi ($n = 470$) (Table 1). These times were chosen to detect early and maintained viral load (Gallardo, personal communication). Time points also coordinated with related studies (Deist et al., 2017a,b, 2018b; Zhang et al., 2018). Production of lachrymal fluid was induced by placing sodium chloride granules on each eye. The resulting fluid accumulation was collected with a pipette. Viral RNA was isolated from the lachrymal fluid using a MagMAX-96 viral RNA isolation kit (Life Technologies, Carlsbad, CA, United States). Isolated RNA was quantified using an LSI VetMAX NDV real-time PCR kit (Life Technologies, Carlsbad, CA, United States) targeted to the matrix protein (M) gene of NDV. Viral RNA was isolated once per sample and quantified via qPCR in duplicate. Mean viral RNA copy number was calculated per sample and log transformed. To test the difference between time points, least squares means were calculated and Student's *t*-test were performed in JMP (SAS Institute, Inc., Cary, NC, United States). In calculating least squares means, effects included qPCR plate, day, room nested within replicate, and sex.

Antibody

Anti-NDV antibody levels in sera were quantified pre-challenge ($n = 453$) and at 10 dpi ($n = 448$) using an IDEXX NDV

TABLE 1 | Descriptive statistics of phenotypes and estimates (SE) of variance components (proportions of phenotypic variance).

| Trait | N ³ | Mean ⁴ | SD ⁵ | Heritability | Dam | Residual |
|---|----------------|-------------------|-----------------|--------------|-------------|-------------|
| Viral load 2 dpi ^{1,2} | 468 | 5.10 | 0.60 | 0.32 (0.1) | – | 0.68 (0.02) |
| Viral load 6 dpi ^{1,2} | 470 | 3.72 | 0.91 | 0.18 (0.1) | – | 0.82 (0.07) |
| Dam antibody ¹ | 139 | 0.53 | 0.22 | – | – | – |
| Antibody pre-challenge ^{1,2} | 453 | –0.84 | 0.60 | 0.26 (0.09) | 0.51 (0.01) | 0.23 (0.00) |
| Antibody 10 dpi ^{1,2} | 448 | –0.06 | 0.26 | 0.24 (0.09) | – | 0.76 (0.01) |
| Growth rate pre-challenge ² | 473 | 10.4 | 1.20 | 0.46 (0.11) | 0.08 (0.04) | 0.47 (0.06) |
| Growth rate post-challenge ² | 470 | 15.4 | 2.76 | 0.21 (0.09) | – | 0.79 (0.3) |

¹Phenotypes log₁₀ transformed. ²Outliers (> 3SD <) removed. ³Number of phenotypic records in the association analysis. ⁴Arithmetic mean. ⁵Standard deviation.

ELISA for chickens (IDEXX Laboratories, Inc., Westbrook, ME, United States) (Table 1). This is the time needed (10 dpi) to generate an acquired immune response (production of specific antibodies). This time also coordinated with related studies (Deist et al., 2017a,b, 2018b; Zhang et al., 2018). Each sample was quantified in duplicate and the average sample:positive (S/P) absorbance ratio was calculated per manufacturer's instructions. To test the difference between time points, a standard least squares effect leverage report and Student's *t*-test were performed in JMP (SAS Institute, Inc., Cary, NC, United States). Effects included day, room nested within replicate, and sex. Antibody levels were also quantified, just prior to the second hatch, on dams (*n* = 139), which had received multiple vaccines against NDV over their lifetime, using the same assay, except plasma was used instead of serum.

Growth Rate

Body weights were recorded in grams on days 0, 13, 21 (0 dpi), 27 (6 dpi), and 31 (10 dpi). Pre-challenge growth rate (*n* = 473) was calculated as grams per day between days 0 and 21. Post-challenge growth rate (*n* = 470) was calculated as grams per day between days 21 and 31.

Genotyping

Whole blood was collected on Whatman FTA cards (Sigma-Aldrich, St. Louis, MO, United States) from all chicks pre-challenge. Genomic DNA was isolated from FTA card punches, dried, and shipped to GeneSeek, Neogen Genomics (Lincoln, NE, United States). DNA was genotyped for 600,000 SNPs using the Axiom Chicken Genotyping Array (Kranis et al., 2013) (Thermo Fisher Scientific, Inc., Waltham, MA, United States). Axiom Chicken Genotyping Array annotation files, release 35, were based on galGal genome version 5.0 (Thermo Fisher Scientific). Quality filtering of genotype data included call rate ≥95 and minor allele frequency ≥0.01. Other filtering metrics (Nclus, FLD, HomRO, HomFLD, HetSO, ConversionType, BB.varX, BB.varY, AB.varX, AB.varY, AA.varX) and requirements are listed in Table 2. These metrics are described in the Axiom Analysis Suite User Guide obtained from Thermo Fisher Scientific (Applied Biosystems, 2017).

Genetic Parameters

Variance components and heritabilities were estimated in ASReml 4 (Gilmour et al., 2015) using the following univariate animal model:

$$Y_{ijk} = \mu + S_i + RR_j + A_k + e_{ijk} \quad (1)$$

where *Y* is the dependent variable of phenotype (viral load 2 and 6 dpi, antibody pre-challenge and 10 dpi, growth rate pre and post-challenge). Sex (*S*) and a combined variable of room and replicate (*RR*) were fitted as fixed effects. Random effects included animal genetic effects (*A*) with a genomic relationship matrix (GRM) computed from SNP genotypes following the procedure described by VanRaden (2008), and residuals (*e*). For viral load at 2 and 6 dpi, qPCR plate was also added as a fixed effect, and for antibody pre-challenge, antibody level of the dam was added as a covariate. The random effect of dam was included for pre-challenge measurements of growth rate and antibody. Phenotypic variance was obtained by summing estimates of variance due to animal, residual, and dam (where applicable). Heritability was calculated as a ratio of the estimates of animal to phenotypic variance.

Association Analysis

Association analyses were performed using the R package GenABEL (Aulchenko, 2015), using a hierarchical generalized linear model (Rönnegård et al., 2010) with the same fixed effects as described for estimation of genetic parameters. The “polygenic_hglm” function was used to fit a polygenic model, with a GRM that was created by GenABEL using the *ibs()* function with the *weight* = “no” option. The “mmscore” function, which is designed to test for association between a trait and genetic polymorphism in samples of related individuals, was used with residuals from *polygenic_hglm* analysis. The *mmscore* function uses the formula

$$\frac{((G - E[G])V^{-1}\text{residualY})^2}{(G - E[G])V^{-1}(G - E[G])} \quad (2)$$

where *G* is the vector of SNP genotypes, *E*[*G*] is a vector of mean genotypic values, *V*^{−1} is the inverse of variance-covariance matrix, and *residualY* are residuals from the trait analysis with *polygenic_hglm*. Together *polygenic_hglm* and *mmscore* function

TABLE 2 | Genotype quality metrics provided by Affymetrix and requirements that were used in quality control filtering.

| Affymetrix genotype metric | Requirement | Brief description of metric ¹ |
|----------------------------|-------------------|---|
| Nclus | $\neq 1$ | Number of genotype clusters |
| Call rate | ≥ 95 | % of samples with a genotype call other than "No Call" |
| MinorAlleleFrequency | ≥ 0.01 | $\min(P_A, P_B)$ |
| FLD | ≥ 3.5 | Measure of the cluster quality of a probeset |
| HomRO | ≥ -0.988 | Distance to zero in the Contrast dimension (X position) from the center of the homozygous cluster that is closest to zero |
| HomFLD | ≥ 10 | Measure of the cluster quality of a probeset for the homozygous genotype clusters |
| HetSO | ≥ -0.21 | Measures how far the heterozygous cluster center sits above the homozygous cluster centers in the Size dimension (Y) |
| ConversionType | $\neq \text{OTV}$ | Probeset classification |
| BB.varX | ≤ 0.85 | Contrast (X position) variance for BB cluster |
| BB.varY | ≤ 0.7 | Size (Y position) variance for BB cluster |
| AB.varX | ≤ 0.75 | Contrast (X position) variance for AB cluster |
| AB.varY | ≤ 0.75 | Size (Y position) variance for AB cluster |
| AA.varX | ≤ 0.79 | Contrast (X position) variance for AA cluster |

¹For detailed description of metrics see *Axiom Analysis Suite User Guide* (Applied Biosystems, 2017).

similarly to the FASTA (Family-based score test for association) method implemented by Chen and Abecasis (2007).

Multiple Test Correction

Genotypes were divided into chromosomes and then further divided into chromosomal segments containing a number of SNPs equal to half the number of animals as described by Waide et al. (2017). The number of independent tests was determined as the sum number of principle components that accounted for 95% of variance between genotypes for each segment (Σn). The number of independent tests was used in a Bonferroni correction to determine 20% suggestive genome-wide thresholds as $0.2/\Sigma n$.

higher than either pre-challenge or at 10 dpi in their chicks.

Growth Rate

Figure 5 shows the population average growth rate pre- and post-challenge and corresponding body weight box plots. Growth rate post-challenge was significantly greater ($P < 0.0001$) than growth rate pre-challenge.

Phenotypic Correlations

Viral load at 2 and 6 dpi were positively correlated (**Table 3**). Pre-challenge antibody level was negatively correlated with both pre and post-challenge growth rate. Post-challenge growth rate

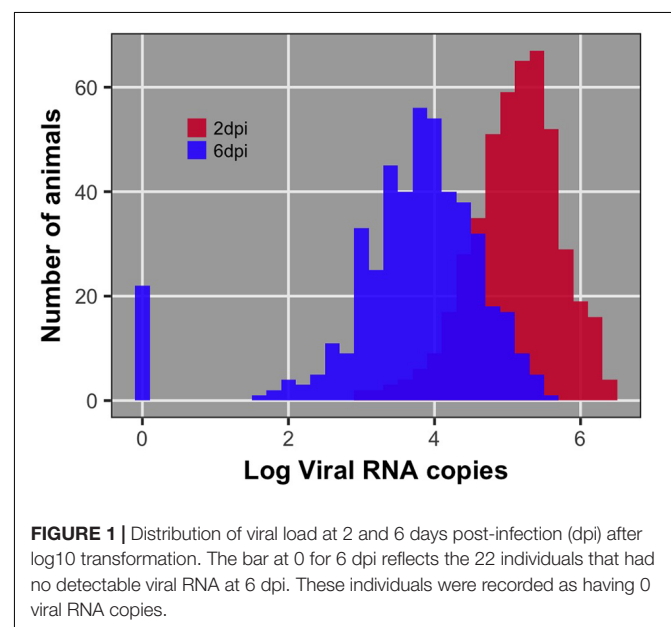
RESULTS

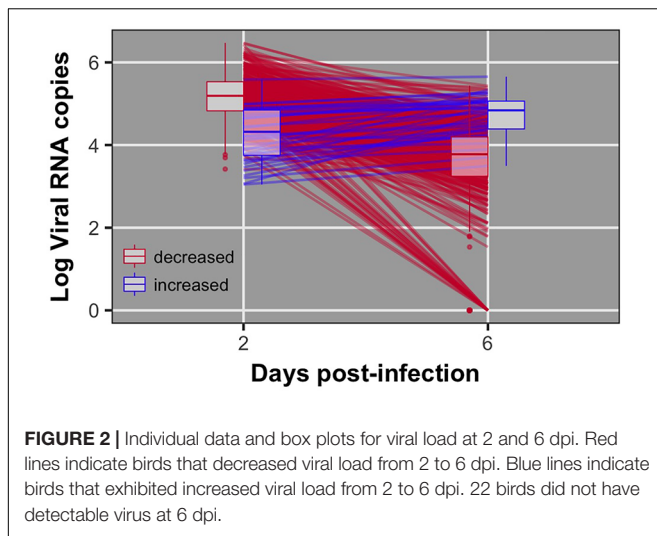
Viral Load

Pre-challenge samples had no measurable virus copies, as expected (data not shown). Distributions of viral load 2 and 6 dpi are shown in **Figure 1**. Viral load was significantly different between 2 and 6 dpi ($P < 0.0001$). Viral load was greater at 2 than 6 dpi for all but 38 birds (8%) (**Figure 2**). By 6 dpi, 22 birds fell below our limit of detection for measurable viral RNA, indicative of viral clearance.

Antibody

Distributions of dam, chick pre-challenge, and chick 10 dpi antibody are shown in **Figure 3**. Pre-challenge anti-NDV antibody levels were measurable but significantly lower than antibody levels at 10 dpi ($P < 0.0001$) for all but 19 birds (4%) (**Figure 4**). These 19 birds were excluded from association analysis for both antibody time points. Antibody levels measured in dams were significantly



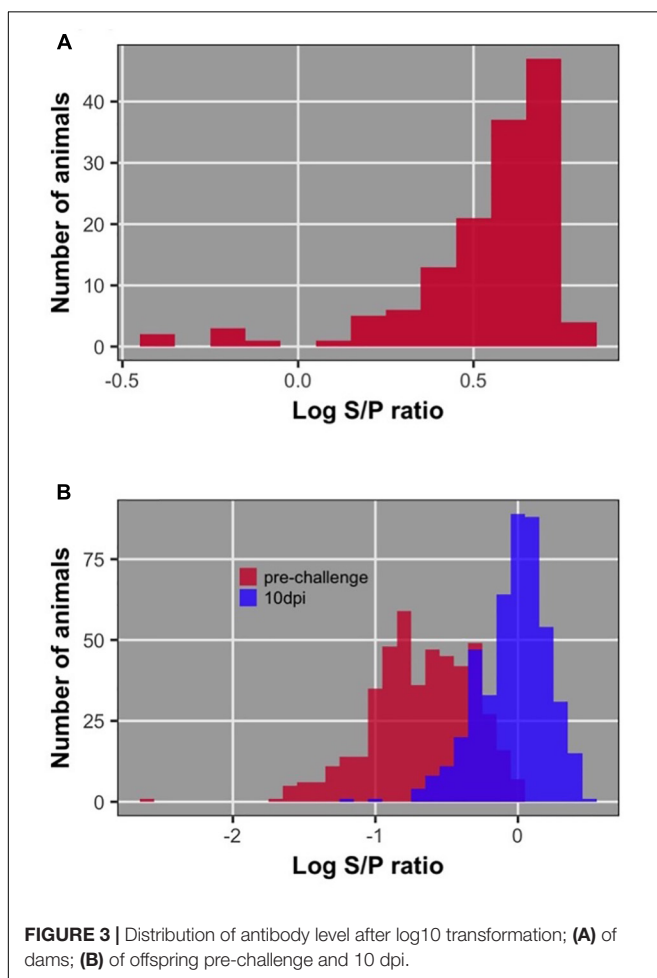


Heritabilities

Heritabilities estimated using AsReml4 were moderate (0.18 to 0.32) for viral load (**Table 1**). Estimates of heritability for pre- and post-challenge antibody levels were similar, 0.26 and 0.24, respectively. Estimates of heritability for Pre- and post-challenge growth rate were moderate, 0.46 and 0.21, respectively.

Genetic Correlations

The estimate of the genetic correlation between viral load at 2 and 6 dpi was high, 0.74 ± 0.21 (**Table 3**). Viral load at 6 dpi and antibody at 10 dpi were negatively correlated (-0.39 ± 0.33); birds with more antibodies had lower viral load. Most pathogen challenge-related traits, with the exception of viral load 6 dpi, were negatively correlated with growth rate pre- and post-challenge (-0.30 to -0.72). The two measures of growth rate had a high positive genetic correlation of 0.72. Standard errors for genetic correlation estimates were moderate, leading some estimates to not differ from 0.



was negatively correlated with viral load at 2 dpi and antibody level pre-challenge, but positively correlated with pre-challenge growth rate.

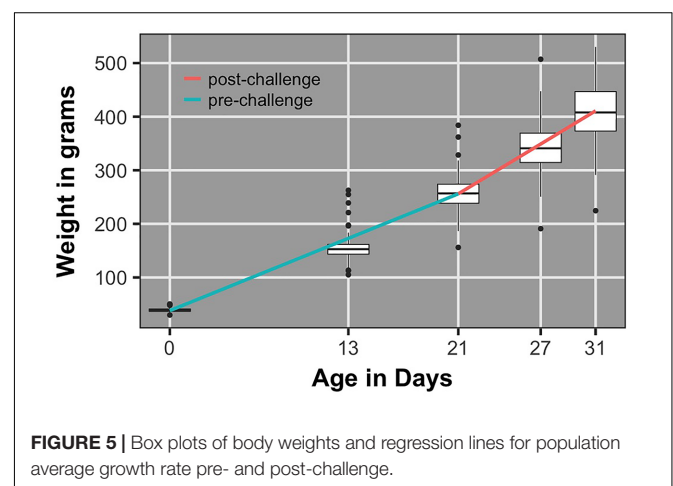
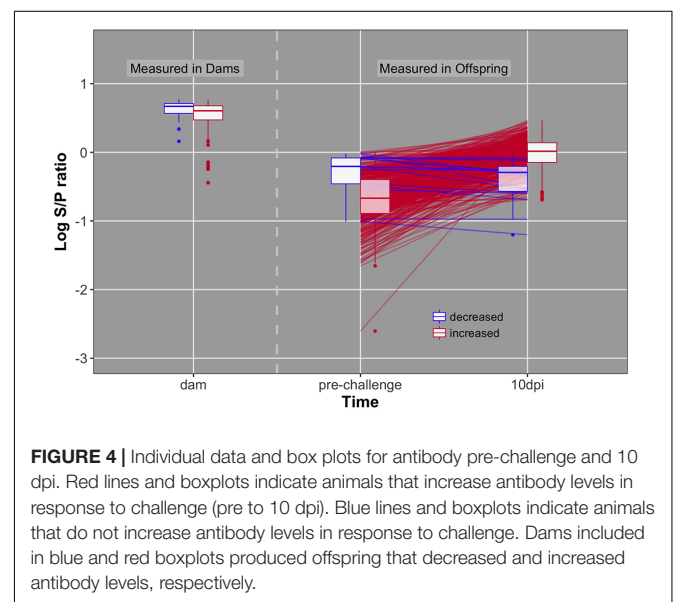


TABLE 3 | Estimates (SE) of phenotypic (above diagonal) and genetic (below diagonal) correlations based on bivariate analyses.

| Trait | VL 2 dpi | VL 6 dpi | AB pre | AB 10 dpi | GR pre | GR post |
|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Viral load 2 dpi | – | 0.18 (0.06) | 0.04 (0.05) | 0.01 (0.05) | –0.05 (0.06) | –0.16 (0.05) |
| Viral load 6 dpi | 0.74 (0.21) | – | –0.03 (0.05) | –0.06 (0.05) | –0.02 (0.06) | –0.05 (0.05) |
| Antibody pre-challenge | 0.12 (0.28) | –0.25 (0.34) | – | 0.06 (0.05) | –0.22 (0.05) | –0.12 (0.05) |
| Antibody 10 dpi | 0.13 (0.31) | –0.39 (0.33) | 0.33 (0.34) | – | –0.05 (0.05) | –0.07 (0.05) |
| Growth rate pre-challenge | –0.11 (0.22) | 0.23 (0.32) | –0.72 (0.19) | –0.42 (0.29) | – | 0.58 (0.04) |
| Growth rate post-challenge | –0.30 (0.26) | 0.21 (0.37) | –0.61 (0.28) | –0.45 (0.33) | 0.72 (0.14) | – |

Alternative Phenotypes

Several alternative phenotypes generated by combination and/or manipulation of individual phenotypes collected in this study were explored: viral load and antibody change over time (difference between time points), viral load clearance (difference between time points divided by 2 dpi level), regression of viral load and antibody measurements over time. However, none were more heritable than the individual phenotypes and most did not have heritability different from 0. Thus, they were not included further in this study.

Association Analysis

After quality control, 476 animals and 340,527 SNPs remained for association analysis. Principle component analysis determined that 44,364 components accounted for 95% of variance between SNPs. Using 44,364 as the number of independent tests and applying Bonferroni correction, the 20% genome-wide significance threshold was 4.508×10^{-6} and used to declare suggestive associations.

Manhattan plots for viral load at 2 and 6 dpi are in **Figures 6, 7**, respectively. One SNP on chromosome 4 was associated with viral load at 6 dpi, while none were associated with viral load 2 dpi (**Table 4**). Association analysis results are reported for antibody pre-challenge and 10 dpi, excluding the 19 birds that did not increase antibody in response to NDV challenge. Manhattan plots for antibody level pre-challenge and at 10 dpi are in **Figures 8, 9**, respectively. Three SNPs were

associated with antibody level pre-challenge, while one SNP was associated with antibody level at 10 dpi (**Table 4**). Two SNPs were associated with growth rate pre-challenge, while none were associated with growth rate post-challenge (**Figures 10, 11** and **Table 4**).

DISCUSSION

Genetic Parameters

Heritabilities for all traits were estimated to be moderate to high, ranging from 0.18 for viral load at 6 dpi to 0.46 for growth rate pre-challenge. Our heritability estimates for antibody levels at 10 dpi are in line with those reported by (Lwelamira et al., 2009) in two Tanzanian chicken ecotypes measured just prior to and 2 weeks post-vaccination (0.27 and 0.29). Peleg et al. (1976) estimated heritability of antibody response to attenuated NDV at 12 dpi to be 0.31 based on the sire variance components. To our knowledge, ours is the first report of heritability for viral load of NDV and growth rate in layer-type birds. The moderate to high heritabilities estimated in this study indicate that all investigated traits can be influenced by selective breeding. Therefore, the means for these traits can be changed over generations.

Negative genetic correlations between pathogen response traits and growth rates indicate that selection for decreased viral load at 2 dpi and for decreased antibody levels is expected to increase pre- and post-challenge growth rate. Many studies have found immune response traits and production/growth traits under challenge to be negatively genetically correlated (Gross et al., 2002; Lwelamira et al., 2009; Hess et al., 2016). Given this information, we can speculate that higher antibody levels, which are often viewed as favorable, may be unfavorable when the desired outcome is to increase disease tolerance. Tolerance is defined as the ability of a host to limit the negative impact of infection (viral in this case) on performance (Bishop, 2012). Tolerance is a good goal for NDV in low income countries, where the virus is relatively ubiquitous and the majority of animals will be infected by the virus at some point in their life. Furthermore, it has been suggested that host tolerance places less pressure on the virus to evolve (Råberg et al., 2009). It must be recognized, however, that standard errors for genetic correlation estimates were moderate, leading some estimates to not differ from 0. A larger sample size will be needed to determine the true significance of genetic correlations.

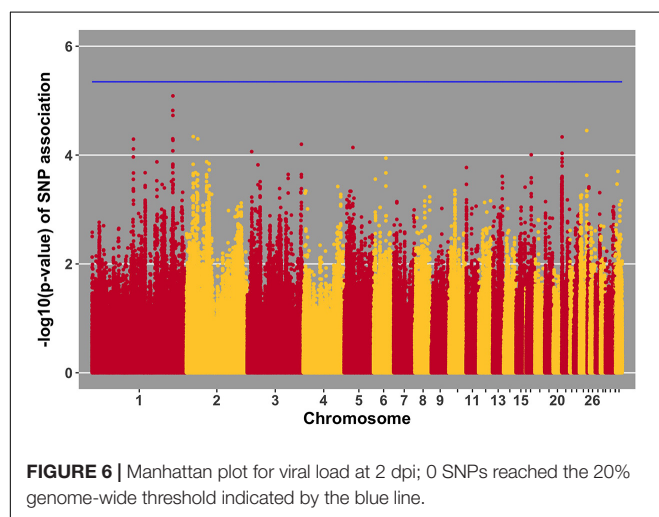
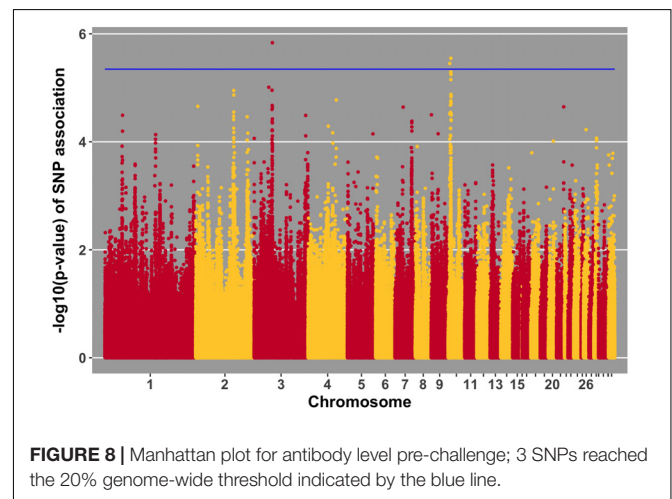
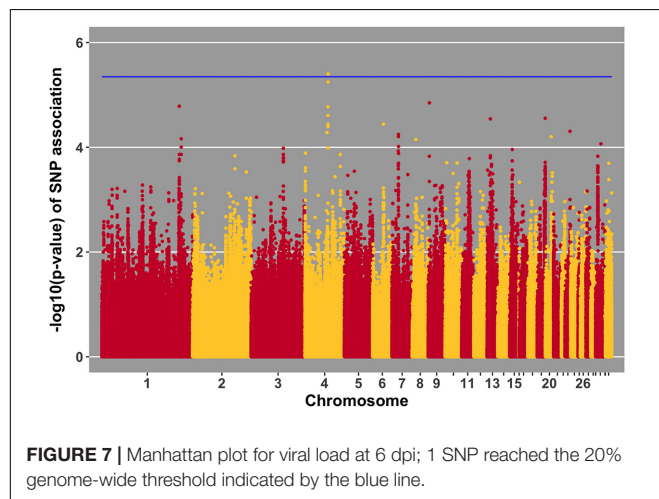


TABLE 4 | Single nucleotide polymorphisms associated with traits, positional candidate genes, and previously reported QTL.

| Trait | SNP | Position ¹ | P-value | Positional Candidate Genes and Location ² | Previous relevant QTL associations |
|------------------------|--------------------------|-----------------------|----------|---|---|
| Antibody pre-challenge | AX-76468260 | 3:38180614 | 1.31E−06 | <i>B3GALNT2</i> , intron <i>GPR137B</i> , upstream 213694 <i>NTPCR</i> , downstream 954041 | Antibody titer to LPS antigen ⁵ |
| | AX-75608938 ⁴ | 10:4095431 | 2.57E−06 | <i>LACTB</i> , intron | Antibody titer to LTA antigen ⁵ |
| | AX-75605132 | 10:2826316 | 3.21E−06 | <i>LINGO1</i> , downstream 58378 <i>HMG20A</i> , downstream 66323 | |
| Antibody 10 dpi | AX-76244799 | 21:3996542 | 4.00E−06 | <i>TARDBP</i> , upstream 65501 <i>APITD1</i> , downstream 267277 <i>CASZ1</i> , upstream 51894 | None related to pathogen response |
| Viral Load 6 dpi | AX-76683655 | 4:53179704 | 4.07E−06 | <i>FAT4</i> , upstream 191754 <i>ANKRD50</i> , upstream 38776 <i>SPRY1</i> , downstream 534821 <i>SULT1B1</i> , upstream 1040403 | Marek's disease-related traits ⁶ |
| | AX-75623995 | 10:8913444 | 3.29E−06 | <i>MAPK6</i> , synon ³ | Carcass weight ⁷ |
| | AX-80773317 ⁴ | 2:27242600 | 4.15E−06 | Novel lincRNA, downstream 384292 Novel lincRNA, downstream 82294 <i>SCIN</i> , downstream 399726 <i>ER81</i> , downstream 85619 | Body weight ^{8,9,10} |
| | | | | | |

¹Chromosome:base pair. ²Location of positional candidate gene (bp from the SNP). ³SNP is within a coding sequence but does not result in a residue change. ⁴SNP is fixed for alternate alleles in Fayoumi and Leghorn inbred lines. ⁵Siwek et al. (2006). ⁶Xu et al. (1998). ⁷Nassar et al. (2012). ⁸Siwek et al. (2004). ⁹Tatsuda and Fujinaka (2001). ¹⁰Uemoto et al. (2009).



The genetic correlation between pre- and post-challenge growth and viral load at 6 dpi was positive, which does not fit the previously mentioned negative trend between pathogen response traits and growth rates, although SE estimates were large. The resource allocation argument may provide an explanation in this case (Gross et al., 2002; Rauw, 2012). Birds that have higher viral load at 6 dpi also have higher pre- and post-challenge growth rates because they use more of their available resources to grow as opposed to clearing the virus.

Viral Load

The 38 birds (8%) that increased viral load from 2 to 6 dpi represent a different kinetic profile of viral clearance than the rest

of the population. Although these 38 birds exhibited a different pattern of viral clearance, there was no evidence for lack of infection or interference of response to challenge. There is no evidence that these birds were less challenged, as all 38 had measurable viral load at 2 dpi, indicating they were infected with NDV. Furthermore, none of the 38 birds were half- or full-sibs to the 19 birds that did not produce antibody in response to challenge. Viral load heritability estimates were not increased by excluding these 38 birds. Therefore, these birds were not excluded from any analyses.

No SNPs reached the suggestive threshold for viral load at 2 dpi, while one SNP reached that threshold for viral load at 6 dpi. For this SNP, located on chromosome 4 at

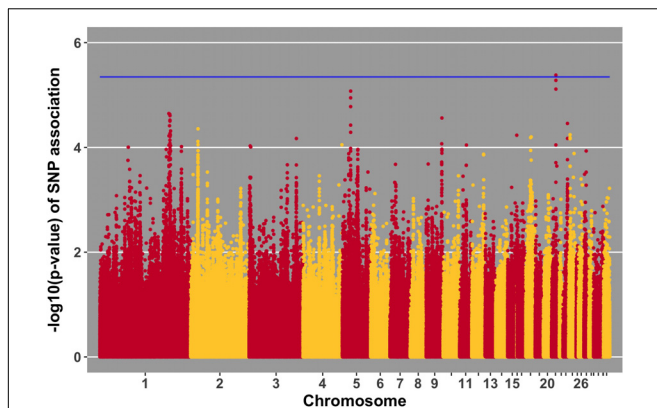


FIGURE 9 | Manhattan plot for antibody 10 dpi, 1 SNP reaches the 20% genome-wide threshold indicated by the blue line.

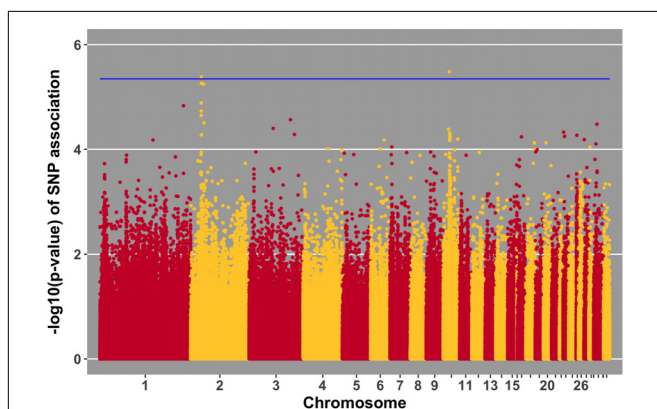


FIGURE 10 | Manhattan plot for growth rate pre-challenge; 2 SNPs reached the 20% genome-wide threshold indicated by the blue line.

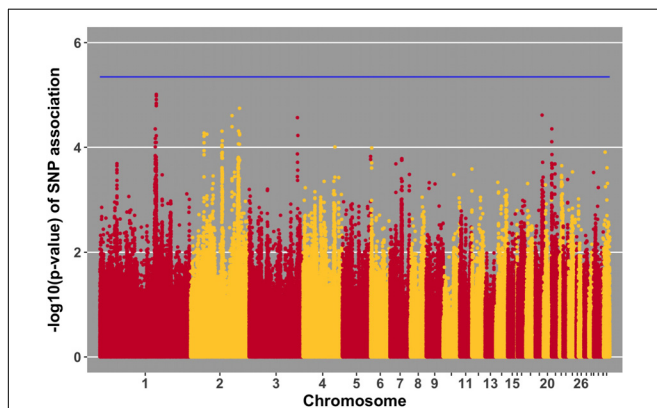


FIGURE 11 | Manhattan plot for growth rate post-challenge, 0 SNPs reach the 20% genome-wide threshold indicated by the blue line.

53 Mb, four genes were located within 1 Mb. This QTL was previously identified in association with Marek's disease-related traits (Xu et al., 1998). The closest gene, *ANKRD50*, was previously found to be down-regulated in tracheal epithelial

cells of an inbred research line of Fayoumi chickens at 2 dpi with NDV compared to non-infected birds (Deist et al., 2017b). Chickens from Fayoumi and Leghorn inbred lines were used in a companion study that had the same experimental design, used the same virus, and measured the same phenotypes as the current study. Deist analyzed transcriptome responses of trachea, lung, and Harderian gland to NDV challenge (Deist et al., 2017a,b, 2018b). Zhang reported transcriptomic changes in the spleen (Zhang et al., 2018). The Fayoumi and Leghorn lines are highly inbred (Fleming et al., 2016) and their responses to various pathogens, including velogenic NDV (Lakshmanan et al., 1996; Cheeseman et al., 2007; Kim et al., 2008; Wang et al., 2014, Deist, 2018a), demonstrate the Fayoumi and Leghorn lines to represent relatively resistant and susceptible genetic research models, respectively. *ANKRD50* functions in endosome to plasma membrane transport (Kvainickas et al., 2017). This is the first reported association of *ANKRD50* with viral infection.

Antibody

It was not expected to have detectable pre-challenge antibody at 20 days of age, because many reports have demonstrated clearance of maternally transferred antibody by this age (Rose and Orlans, 1981; Liu and Higgins, 1990; Grindstaff et al., 2003; Hamal et al., 2006). However, the dams of challenged chicks were 'hyperimmunized,' as they had received 5 immunizations for NDV prior to production of the chicks used in this study. Thus, we believe that the passive maternal antibody still circulating at 20 days of age may have interfered with the response to NDV challenge, specifically in the 19 chicks that did not increase level of antibody between pre-challenge and 10 dpi. Maternal antibody interference with vaccine response is a known phenomenon (Richey and Schmittle, 1962; Eidson et al., 1976). Because these 19 chicks were likely unable to respond to the vaccine appropriately because of passive antibody interference, we conducted analyses both with and without these chicks included. Heritability of pre-challenge antibody increased from 0.20 to 0.26 with exclusion of these birds. The same trend of increasing heritability was seen for antibody 10 dpi, from 0.19 to 0.24. Three suggestive QTL were found when excluding these 19 birds, while only two of the three were found when using the full dataset. These analyses provide evidence that passive antibody interference caused "noise" in the antibody response data; therefore these 19 birds were excluded from the association analysis for antibody pre-challenge and at 10 dpi. We expect that dams in low income countries would also have relatively high amounts of anti-NDV antibodies due to high environmental levels of NDV and repeated exposure to the virus.

Pre-challenge antibody did not differ significantly between the three replicates suggesting that maternal antibody transfer level did not differ significantly due to the time between the three hatches. Dam antibodies were measured from plasma, while chick antibodies were measured from serum. Previous studies have shown that antibody measured in plasma and serum are highly correlated (Cherpes et al., 2003; Siev et al., 2011), suggesting the validity of comparing levels of antibody between dams' plasma and chicks' serum in the current study.

Three SNPs, in two QTL, were suggestively associated with antibody level pre-challenge. The strongest association was on chromosome 3 at 38.2 Mb. This SNP was within the intron of *B3GALNT2*. *B3GALNT2* was previously found to be more highly expressed in the Harderian gland of Fayoumis compared to Leghorns at 2 days post-NDV inoculation (Deist et al., 2018b). *B3GALNT2* functions in protein glycosylation (Stevens et al., 2013). We present a novel association of *B3GALNT2* with viral infection. One gene, *GPR137B*, was near the SNP, 213,694 bp upstream. *GPR137B* is a lysosomal integral membrane protein predicted to function in signal transduction (Gao et al., 2012). The QTL on chromosome 3 for antibody level pre-challenge, was previously associated with antibody titer to LPS antigen (Siwek et al., 2006).

The second QTL for antibody pre-challenge on chromosome 10 contained two SNPs. The strongest SNP within the chromosome 10 QTL was within the intron of the *LACTB* gene. This SNP was fixed for alternate alleles in the Fayoumi and Leghorn lines, evaluated by 600k Axiom Chicken Genotyping Array data from 10 birds per line. *LACTB* promotes intra-mitochondrial membrane organization through polymerization (Polianskyte et al., 2009). This is the first identified association of *LACTB* with antibody production.

The second SNP within the chromosome 10 QTL was near two genes, *LINGO1* and *HMG20A*. *LINGO1* was previously found to be down-regulated in the lung of Fayoumi chickens at 2 dpi with NDV, compared to non-infected birds (Deist et al., 2017a). *LINGO1* was also less expressed in the Harderian gland of Fayoumis compared to Leghorns at 2 days after challenge with NDV (Deist et al., 2018b). When comparing the expression in the lung of non-challenged birds, Fayoumi chickens expressed more *LINGO1* than Leghorns (Deist et al., 2017a). *LINGO1* is a transmembrane protein functioning in signal transduction (Mi et al., 2004). *HMG20A* exhibited more expression in the lung of non-challenged Leghorn chickens compared to Fayoumi chickens (Deist et al., 2017a). In tracheal epithelial cells at 2 and 10 days post-NDV infection, Leghorns expressed more *HMG20A* than Fayoumis (Deist et al., 2017b). *HMG20A* has been shown to bind to viral DNA *in vitro* (Hsiao et al., 2006). The second antibody pre-challenge QTL on chromosome 10 was previously associated with antibody titer to LTA antigen (Siwek et al., 2006).

Antibody level at 10 dpi was associated with one SNP, located at 3.9 Mb on chromosome 21. Three genes were nearby, *TARDBP*, *APITD1*, and *CASZ1*. *CASZ1* was down-regulated 2 days post-NDV challenge in tracheal epithelial cells of Leghorn chickens compared to non-challenged birds (Deist et al., 2017b). *TARDBP* functions in negative regulation by host of viral transcription (GO biological process) and was previously implicated as part of the influenza-host interactome using human and mammalian cell lines *in vitro* (Heaton et al., 2016).

Growth Rate

Compared to the management guide for the Hy-Line Brown commercial layers, our birds had higher body weights across

all weeks partially due to the inclusion of male chicks in our experimental population (Hy-Line International, 2016). However, the growth rate trajectories between our birds and the management guide are roughly parallel, suggesting we are not seeing a large depression due to challenge.

Growth rate pre-challenge was associated with two SNPs on chromosomes 10 and 2. The SNP on chromosome 10 was within the *MAPK6* gene, which functions in phosphorylation. This QTL co-localizes with a previous association for carcass weight (Nassar et al., 2012).

The *ER81* gene, near the SNP for pre-challenge growth rate on chromosome 2, functions in transcription regulation. This QTL has been previously been identified to be associated with body weight in three independent populations (Tatsuda and Fujinaka, 2001; Siwek et al., 2004; Uemoto et al., 2009).

Support of Expression Studies for Suggestive SNP Associations

Incorporating previous gene expression data can improve the value of GWAS data, especially when significant expression data coincides with suggestive (near-significant) SNPs (Cheng et al., 2013). The SNP on chromosome 5 with the lowest *p*-value of association with viral load at 2 dpi (Figure 6), is within a gene (*PAMR1*) that was previously found to be differentially expressed in tracheal epithelial cells at 2 and 6 days post-NDV infection (Deist et al., 2017b). At both time points, Leghorn chickens expressed higher levels of *PAMR1* compared to Fayoumis. The Leghorn chickens in the referenced study had significantly more viral genome transcripts in the trachea at 2 dpi. Perhaps the difference in viral load between the two lines is due in part to expression differences in this gene and provide support for the near significant GWAS results.

We identified a suggestive QTL on chromosome 4 at 53 Mb for viral load at 6 dpi (Figure 7). Several SNPs in the location of the QTL fell just below the threshold. One of these SNPs (*p*-value of association 5.22×10^{-5}) is within the *ADAMTS3* gene. In samples from the birds utilized in this GWAS study, the *ADAMTS3* gene was shown to be down-regulated in the spleen 6 days after NDV challenge (Zhang et al., 2018). Integration of this information provides further evidence for the existence of the identified QTL for viral load at 6 dpi on chromosome 4.

The *OFD1* gene encompasses three SNPs within the near-significant QTL on chromosome 1 for growth rate post-challenge (Figure 11). *OFD1* functions in primary cilium organization and assembly (Ferrante et al., 2006). This gene was shown to exhibit lower expression in the Harderian gland of Leghorns compared to Fayoumis, 2 days after NDV challenge (Deist et al., 2018b). Perhaps the differential expression of *OFD1* contributes to the susceptible/resistant phenotypes of the Leghorn/Fayoumi lines. Overall, *OFD1* may play a role in NDV tolerance – performance (growth) under challenge.

CONCLUSION

Six suggestive QTL associated with response to NDV and/or growth were identified. Some were novel and others confirmed previously reported associations with related traits. Additionally, previous RNA-seq analysis provided support for several of the genes located in or near the QTL of the current study. Considering the trend of negative genetic correlation between antibody and Newcastle Disease tolerance (growth under disease) and estimates of moderate to high heritability, we provide evidence that these NDV response traits can be influenced through selective breeding. This information can inform breeding decisions for the production of chickens that will be raised in NDV endemic areas once more knowledge of the relationship of antibody and viral load with mortality is obtained. Producing chickens that perform favorably in challenging environments will ultimately increase the supply of quality protein for human consumption.

AVAILABILITY OF DATA AND MATERIALS

The data that support the findings of this study are available from Hy-Line International but restrictions apply to the availability of these data, which were used under license for the current study, and thus are not publicly available. However, data are available from the authors upon reasonable request to SL and with permission of Hy-Line International.

REFERENCES

- Applied Biosystems (2017). *Axiom Analysis Suite 3.1 User Guide*. Available at: <https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/axiom-analysis-suite.html>
- Aulchenko, Y. (2015). *Package GenABEL - R Package Reference Manual*. 143. Available at: <https://cran.r-project.org/web/packages/GenABEL/index.html>
- Bishop, S. C. (2012). A consideration of resistance and tolerance for ruminant nematode infections. *Front. Genet.* 3:168. doi: 10.3389/fgene.2012.00168
- Cheeseman, J. H., Kaiser, M. G., Ciraci, C., Kaiser, P., and Lamont, S. J. (2007). Breed effect on early cytokine mRNA expression in spleen and cecum of chickens with and without *Salmonella enteritidis* infection. *Dev. Comp. Immunol.* 31, 52–60. doi: 10.1016/j.dci.2006.04.001
- Chen, W. M., and Abecasis, G. R. (2007). Family-based association tests for genome-wide association scans. *Am. J. Hum. Genet.* 81, 913–926. doi: 10.1086/521580
- Cheng, H. H., Kaiser, P., and Lamont, S. J. (2013). Integrated genomic approaches to enhance genetic resistance in chickens. *Annu. Rev. Anim. Biosci.* 1, 239–260. doi: 10.1146/annurev-animal-031412-103701
- Cherpes, T. L., Meyn, L. A., and Hillier, S. L. (2003). Plasma versus serum for detection of herpes simplex virus type 2-specific immunoglobulin G antibodies with a glycoprotein G2-based enzyme immunoassay. *J. Clin. Microbiol.* 41, 2758–2759. doi: 10.1128/JCM.41.6.2758
- Cole, R. K., and Hutt, F. B. (1961). Genetic differences in resistance to Newcastle disease. *Avian. Dis.* 5, 205–214. doi: 10.2307/1587625
- Deist, M. S., Gallardo, R. A., Bunn, D. A., Dekkers, J. C. M., Zhou, H., and Lamont, S. J. (2017a). Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. *BMC Genomics* 18:989. doi: 10.1186/s12864-017-4380-4

AUTHOR CONTRIBUTIONS

HZ, RG, JD, and SL designed the study. KR, AW, JD, and SL performed the animal experiments and collected phenotypes. KR performed the lab work related to quantifying phenotypes and DNA extraction. KR performed the data analysis with inputs from AW, JD, HZ, and SL. KR wrote the initial draft of the manuscript. KR, AW, RG, TK, HZ, JD, and SL provided critical revision. All authors read and approved the final manuscript.

FUNDING

KR was supported by a USDA National Needs Fellowship (2013-38420-20496). This work was supported by USAID Feed the Future Innovation Lab for the Genomics to Improve Poultry and Hatch project #5357. This study is made possible by the generous support of the American people through USAID. The contents are the responsibility of the Feed the Future Innovation Lab for Genomics to Improve Poultry and do not necessarily reflect the views of USAID or the United States Government.

ACKNOWLEDGMENTS

We thank Hy-Line International for providing the birds used in this experiment as well as the Lamont lab members, especially Michael Kaiser, for their help in organization and sample collection.

- Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., Dekkers, J. C. M., Zhou, H., et al. (2017b). Novel mechanisms revealed in the trachea transcriptome of resistant and susceptible chicken lines following infection with Newcastle disease virus. *Clin. Vaccine Immunol.* 24, 1–17. doi: 10.1128/CVI.00027-17
- Deist, M. (2018a). *Host Transcriptomic Response to Newcastle Disease Virus in Relatively Resistant and Susceptible Inbred Chicken Lines*. Doctoral dissertation, Iowa State University, Ames, IA.
- Deist, M. S., Gallardo, R. A., Bunn, D. A., Kelly, T. R., Dekkers, J. C. M., Zhou, H., et al. (2018b). Novel analysis of the Harderian gland transcriptome response to Newcastle disease virus in two inbred chicken lines. *Sci. Rep.* 8:6558. doi: 10.1038/s41598-018-24830-0
- Eidson, C. S., Kleven, S. H., and Villegas, P. (1976). Efficacy of intratracheal administration of Newcastle disease vaccine in day-old chicks. *Poult. Sci.* 44, 1252–1267. doi: 10.3382/ps.0551252
- Ferrante, M. I., Zullo, A., Barra, A., Bimonte, S., Messaddeq, N., Studer, M., et al. (2006). Oral-facial-digital type I protein is required for primary cilia formation and left-right axis specification. *Nat. Genet.* 38, 112–117. doi: 10.1038/ng1684
- Fleming, D. S., Koltes, J. E., Fritz-Waters, E. R., Rothschild, M. F., Schmidt, C. J., Ashwell, C. M., et al. (2016). Single nucleotide variant discovery of highly inbred Leghorn and Fayoumi chicken breeds using pooled whole genome resequencing data reveals insights into phenotype differences. *BMC Genomics* 17:812. doi: 10.1186/s12864-016-3147-7
- Fulton, J. E. (2012). Genomic selection for poultry breeding. *Anim. Front.* 2, 30–36. doi: 10.2527/af.2011-0028
- Gao, J., Xia, L., Lu, M., Zhang, B., Chen, Y., Xu, R., et al. (2012). TM7SF1 (GPR137B): a novel lysosome integral membrane protein. *Mol. Biol. Rep.* 39, 8883–8889. doi: 10.1007/s11033-012-1755-0

- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml User Guide Release 4.1 Functional Specification*. Available at: www.vsnl.co.uk
- Gordon, C. D., Beard, C. W., Hopkins, S. R., and Siegel, H. S. (1970). Chick mortality as a criterion for selection toward resistance or susceptibility to Newcastle disease. *Poult. Sci.* 50, 783–789. doi: 10.3382/ps.0500783
- Grindstaff, J. L., Iii, E. D. B., and Ketterson, E. D. (2003). Immune function across generations: integrating mechanism and evolutionary process in maternal antibody transmission. *Proc. R. Soc. Lond. B Biol. Sci.* 270, 2309–2319. doi: 10.1098/rspb.2003.2485
- Gross, W. B., Siegel, P. B., and Pierson, F. W. (2002). Effects of genetic selection for high or low antibody response on resistance to a variety of disease challenges and the relationship of resource allocation. *Avian Dis.* 46, 1007–1010. doi: 10.1637/0005-2086(2002)046[1007:EOGSFH]2.0.CO;2
- Guèye, E. F. (2000). The role of family poultry in poverty alleviation, food security and the promotion of gender equality in rural Africa. *Outlook Agric.* 29, 129–136. doi: 10.5367/0000000000101293130
- Hamal, K. R., Burgess, S. C., Pevzner, I. Y., and Erf, G. F. (2006). Maternal antibody transfer from dams to their egg yolks, egg whites, and chicks in meat lines of chickens. *Poult. Sci.* 85, 1364–1372. doi: 10.1093/ps/85.8.1364
- Havenstein, G. B., Ferket, P. R., and Qureshi, M. A. (1994). Growth, livability, and feed conversion of 1957 vs 1991 broilers when fed “typical” 1957 and 1991 broiler diets. *Poult. Sci.* 73, 1785–1794. doi: 10.3382/ps.0731785
- Heaton, N. S., Moshkina, N., Fenouil, R., Gardner, T. J., Aguirre, S., Shah, P. S., et al. (2016). Targeting viral proteostasis limits Influenza virus, HIV, and Dengue virus infection. *Immunity* 44, 46–58. doi: 10.1016/j.immuni.2015.12.017
- Hess, A. S., Islam, Z., Hess, M. K., Rowland, R. R. R., Lunney, J. K., Doeschl-Wilson, A., et al. (2016). Comparison of host genetic factors influencing pig response to infection with two North American isolates of porcine reproductive and respiratory syndrome virus. *Genet. Sel. Evol.* 48:43. doi: 10.1186/s12711-016-0222-0
- Hsiao, J. C., Chao, C. C., Young, M. J., Chang, Y. T., Cho, E. C., and Chang, W. (2006). A Poxvirus host range protein, CP77, binds to a cellular protein, HMG20A, and regulates its dissociation from the vaccinia virus genome in CHO-K1 cells. *J. Virol.* 80, 7714–7728. doi: 10.1128/JVI.00207-06
- Hy-Line International (2016). *Hy-Line Brown Management Guide*. Available at: www.hyline.com/aspx/general/dynamicpage.aspx?id=255
- Kim, D. K., Lillehoj, H. S., Hong, Y. H., Park, D. W., Lamont, S. J., Han, J. Y., et al. (2008). Immune-related gene expression in two B-complex disparate genetically inbred Fayoumi chicken lines following *Eimeria maxima* infection. *Poult. Sci.* 87, 433–443. doi: 10.3382/ps.2007-00383
- Kitalyi, A. J. (1998). *Village Chicken Production Systems in Rural Africa: Household Food Security and Gender Issues*. Rome: Food & Agriculture Organization.
- Kranis, A., Gheys, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59. doi: 10.1186/1471-2164-14-59
- Kvainickas, A., Orgaz, A. J., Nägele, H., Diedrich, B., Heesom, K. J., Dengjel, J., et al. (2017). Retromer- and WASH-dependent sorting of nutrient transporters requires a multivalent interaction network with ANKRD50. *J. Cell Sci.* 130, 382–395. doi: 10.1242/jcs.196758
- Lakshmanan, N., Kaiser, M. G., and Lamont, S. J. (1996). “Marek’s disease resistance in MHC congenic lines from Leghorn and Fayoumi breeds,” in *Current Research on Marek’s Disease*, ed. R. F. Silva (Kennett Square, PA: American Association of Avian Pathologists), 57–62.
- Liu, S. S., and Higgins, D. A. (1990). Yolk-sac transmission and post-hatching ontogeny of serum immunoglobulins in the duck (*Anas platyrhynchos*). *Comp. Biochem. Physiol.* 97B, 637–644. doi: 10.1016/0305-0491(90)90100-8
- Lwelamira, J., Kifaro, G. C., and Gwakisa, P. S. (2009). Genetic parameters for body weights, egg traits and antibody response against Newcastle Disease Virus (NDV) vaccine among two Tanzania chicken ecotypes. *Trop. Anim. Health Prod.* 41, 51–59. doi: 10.1007/s11250-008-9153-2
- Mayers, J., Mansfield, K. L., and Brown, I. H. (2017). The role of vaccination in risk mitigation and control of Newcastle disease in poultry. *Vaccine* 35, 5974–5980. doi: 10.1016/j.vaccine.2017.09.008
- Mi, S., Lee, X., Shao, Z., Thill, G., Ji, B., Relton, J., et al. (2004). LINGO-1 is a component of the Nogo-66 receptor/p75 signaling complex. *Nat. Neurosci.* 7, 221–228. doi: 10.1038/nn1188
- Nassar, M. K., Goraga, Z. S., and Brockmann, G. A. (2012). Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: II. Muscle weight and carcass composition. *Anim. Genet.* 43, 739–745. doi: 10.1111/j.1365-2052.2012.02344.x
- Peleg, B. A., Soller, M., Ron, N., Hornstein, K., Brody, T., and Kalmar, E. (1976). Familial differences in antibody response of broiler chickens to vaccination with attenuated and inactivated Newcastle disease virus vaccine. *Avian Dis.* 20, 661–668. doi: 10.2307/1589446
- Pitcovski, J., Heller, D. E., Cahaner, A., and Peleg, B. A. (1987). Selection for early responsiveness of chicks to *Escherichia coli* and Newcastle disease virus. *Poult. Sci.* 66, 1276–1282. doi: 10.3382/ps.0661276
- Polianskyte, Z., Peitsaro, N., Dapkunas, A., Liobikas, J., Soliymani, R., Lalowski, M., et al. (2009). LACTB is a filament-forming protein localized in mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18960–18965. doi: 10.1073/pnas.0906734106
- Räberg, L., Graham, A. L., and Read, A. F. (2009). Decomposing health: tolerance and resistance to parasites in animals. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 37–49. doi: 10.1098/rstb.2008.0184
- Rauw, W. M. (2012). Immune response from a resource allocation perspective. *Front. Genet.* 3:267. doi: 10.3389/fgene.2012.00267
- Richey, D. J., and Schmittle, S. C. (1962). The effect of congenital passive immunity levels on the response of chicks to Newcastle disease vaccination. *J. Immunol.* 89, 344–347.
- Rönnegård, L., Shen, X., and Alam, M. (2010). hglm: a package for fitting hierarchical generalized linear models. *R J.* 2, 20–28.
- Rose, M. E., and Orlans, E. (1981). Immunoglobulins in the egg, embryo and young chick. *Dev. Comp. Immunol.* 5, 15–20. doi: 10.1016/S0145-305X(81)80003-1
- Siev, M., Yu, X., Prados-rosales, R., Martiniuk, F. T., Casadevall, A., and Achkar, J. M. (2011). Correlation between serum and plasma antibody titers to mycobacterial antigens. *Clin. Vaccine Immunol.* 18, 173–175. doi: 10.1128/CI.00325-10
- Siwek, M., Buitenhuis, B., Cornelissen, S., Nieuwland, M., Knol, E. F., Crooijmans, R., et al. (2006). Detection of QTL for innate: non-specific antibody levels binding LPS and LTA in two independent populations of laying hens. *Dev. Comp. Immunol.* 30, 659–666. doi: 10.1016/j.dci.2005.09.004
- Siwek, M., Cornelissen, S. J. B., Buitenhuis, A. J., Nieuwland, M. G. B., Bovenhuis, H., Crooijmans, R. P. M. A., et al. (2004). Quantitative trait loci for body weight in layers differ from quantitative trait loci specific for antibody responses to sheep red blood cells. *Poult. Sci.* 83, 853–859. doi: 10.1093/ps/83.6.853
- Soller, M., Heller, D., Peleg, B., Ron-Kuper, N., and Hornstein, K. (1981). Genetic and phenotypic correlations between immune response to *Escherichia coli* and to Newcastle disease virus vaccines. *Poult. Sci.* 60, 49–53. doi: 10.3382/ps.0600049
- Stevens, E., Carss, K. J., Cirak, S., Foley, A. R., Torelli, S., Willer, T., et al. (2013). Mutations in B3GALNT2 cause congenital muscular dystrophy and hypoglycosylation of α -dystroglycan. *Am. J. Hum. Genet.* 92, 354–365. doi: 10.1016/j.ajhg.2013.01.016
- Tatsuda, K., and Fujinaka, K. (2001). Genetic mapping of the QTL affecting body weight in chickens using a F2 family. *Br. Poult. Sci.* 42, 333–337. doi: 10.1080/00071660120055296
- Uemoto, Y., Sato, S., Odawara, S., Nokata, H., Oyama, Y., Taguchi, Y., et al. (2009). Genetic mapping of quantitative trait loci affecting growth and carcass traits in F2 intercross chickens. *Poult. Sci.* 88, 477–482. doi: 10.3382/ps.2008
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Waide, E. H., Tuggle, C. K., Serão, N. V. L., Schroyen, M., Hess, A., Rowland, R. R. R., et al. (2017). Genome wide association of piglet responses to infection with one of two porcine reproductive and respiratory syndrome virus isolates. *J. Anim. Sci.* 95, 16–38. doi: 10.2527/jas.2016.0874
- Wang, Y., Lupiani, B., Reddy, S. M., Lamont, S. J., and Zhou, H. (2014). RNA-seq analysis revealed novel genes and signaling pathway associated with disease resistance to avian influenza virus infection in chickens. *Poult. Sci.* 93, 485–493. doi: 10.3382/ps.2013-03557
- Xu, S., Yonash, N., Vallejo, R. L., and Cheng, H. H. (1998). Mapping quantitative trait loci for binary traits using a heterogeneous residual variance model: an application to Marek’s disease susceptibility in chickens. *Genetica* 104, 171–178. doi: 10.1023/A:1003522902078

Zhang, J., Kaiser, M. G., Deist, M. S., Gallardo, R. A., David, A. B., Kelly, T. R., et al. (2018). Transcriptome analysis in spleen reveals differential regulation of response to Newcastle disease virus in two chicken lines. *Sci. Rep.* 8, 1–13. doi: 10.1038/s41598-018-19754-8

Conflict of Interest Statement: Hy-Line International made the in-kind contribution of the animals studied. AW is employed by Hy-Line International and Iowa State University. AW helped with the animal experiments and phenotype collections and provided input on data analysis. Hy-Line International is interested in the outcome of this experiment in regards to their commercial product, but this had no influence on the outcomes of the experiment or this manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Rowland, Wolc, Gallardo, Kelly, Zhou, Dekkers and Lamont. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Overexpression of Chicken IRF7 Increased Viral Replication and Programmed Cell Death to the Avian Influenza Virus Infection Through TGF-Beta/FoxO Signaling Axis in DF-1

Tae Hyun Kim^{1,2} and Huaijun Zhou^{1,2*}

¹ Department of Animal Science, University of California, Davis, Davis, CA, United States, ² Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA, United States

OPEN ACCESS

Edited by:

Mark S. Fife,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Irit Davidson,
Kimron Veterinary Institute, Israel
Sascha Trapp,
INRA Centre Val de Loire, France

*Correspondence:

Huaijun Zhou
hzhou@ucdavis.edu

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 30 May 2018

Accepted: 06 September 2018

Published: 25 September 2018

Citation:

Kim TH and Zhou H (2018)
Overexpression of Chicken IRF7
Increased Viral Replication
and Programmed Cell Death to the
Avian Influenza Virus Infection
Through TGF-Beta/FoxO Signaling
Axis in DF-1. *Front. Genet.* 9:415.
doi: 10.3389/fgene.2018.00415

During mammalian viral infections, interferon regulatory factor 7 (IRF7) partners with IRF3 to regulate the type I interferon response. In chickens, however, it is still unclear how IRF7 functions in the host innate immune response, especially given that IRF3 is absent. To further elucidate the functional role of chicken IRF7 during avian influenza virus (AIV) infection, we generated inducible IRF7 overexpression DF-1 cell lines and performed *in vitro* infection using low pathogenic AIVs (LPAIVs). Overexpression of IRF7 resulted in higher viral replication of H6N2 and H10N7 LPAIVs compared to empty vector control cells regardless of IRF7 expression level. In addition, a high rate of induced cell death was observed due to elevated level of IRF7 upon viral infection. RNA-seq and subsequent transcriptome analysis of IRF7 overexpression and control cells discovered candidate genes possibly controlled by chicken IRF7. Functional annotation revealed potential pathways modulated by IRF7 such as TGF-beta signaling pathway, FoxO signaling pathway and cell structural integrity related pathways. Next, we analyzed the host response alteration due to the IRF7 overexpression and additionally discovered the possible connection of chicken IRF7 and JAK-STAT signaling pathway. These findings suggest that chicken IRF7 could modulate a wide range of cellular processes in the host innate immune response thus meticulous control of IRF7 expression is crucial to the host in response to AIV infection.

Keywords: AIV, avian influenza, chicken, DF-1 cell line, IRF7, overexpression, RNA-seq

INTRODUCTION

Avian influenza virus (AIV) is one of the major pathogens that significantly impacts the poultry industry worldwide (Olsen et al., 2006; Swayne, 2012). For example, recent high pathogenic avian influenza (HPAI) outbreaks between late-2014 to mid-2015 in the United States resulted in the death of more than 50 million birds, an estimated 12% of the layer chickens and eight percent of the meat turkeys raised in the United States that year (Jhung and Nelson, 2015; Ramos et al., 2017). Declined production and HPAI-related trade restrictions further contributed to the significant economic loss to the industry (Ramos et al., 2017). Current strategies for controlling AIV primarily rely on passive measures such as quarantine and slaughter, partially due to our

limited understanding of the chicken antiviral response compared to mammals (Goossens et al., 2013). A better understanding of the host antiviral response in chickens could provide critical information to develop improved prevention strategies as well as novel therapeutics against AIV (Downing et al., 2009; Magor et al., 2013).

Interferons (IFNs) are known to trigger host innate immune responses against viral infection by activating signal transduction pathways (Der et al., 1998; de Weerd et al., 2007), and currently more than 3,800 IFN regulated genes have been reported according to the Interferome database (Rusinova et al., 2013). IRF7 is well known as the master transcription factor that interacts with IRF3 to initiate the type I IFN response in mammals (Honda et al., 2005; Honda and Taniguchi, 2006). Even though chickens can induce type I IFNs robustly in response to viral infection (Der et al., 1998), avian species lack IRF3, and the precise molecular and cellular mechanisms of IFN regulation remain to be elucidated (Huang et al., 2010; Santhakumar et al., 2017). Based on this, we hypothesized that the chicken IRF7 may have a conserved function in regulating antiviral response in chickens, yet the signaling cascade and mechanism of action could be species-specific.

In our recent analysis of stable overexpression and knockdown of IRF7 in chicken DF-1 cell lines followed by mimicking viral infection with dsRNA analog poly(I:C), we demonstrated that the primary function of IRF7 as type I IFN regulator may be conserved (Kim and Zhou, 2015). Constitutive overexpression of IRF7 resulted in upregulation of *IFNB* upon poly(I:C) induction whereas IRF7 knockdown caused downregulation of *IFNA* (Kim and Zhou, 2015). Further transcriptome analysis revealed more than 60 novel candidate genes that are potentially regulated by IRF7, suggesting a distinct function of chicken IRF7 (Kim and Zhou, 2015). Another study demonstrated that the knockdown of IRF7 by siRNA limited *IFNA*, *IFNB*, and *STAT1* mRNA expression and increased Newcastle disease virus replication in chicken embryonic fibroblasts (CEFs), suggesting the functional role of IRF7 as a type I IFN regulator (Wang Y. et al., 2014).

To further elucidate the functional role of chicken IRF7 in the context of AIV infection, we took advantage of the inducible expression system to control the expression level of IRF7 in DF-1 cells and infected the established cell lines with two low pathogenic AIV (LPAIV) strains. Correlation between the IRF7 expression level and the AIV replication phenotype was investigated with different levels of IRF7 induction. In addition, we analyzed the transcriptome of IRF7 overexpression and control cells by RNA-seq after LPAIV or mock infection to examine candidate genes and pathways that are potentially modulated by IRF7 upon AIV infection.

MATERIALS AND METHODS

Expression Plasmid Construction

Chicken *IRF7* coding sequence (CDS, KP_096419) was cloned into the piggyBac(pB) cumate expression inducible plasmid (System Biosciences, Mountain View, CA, United States) which controls the expression level by cumate gene switch

(pB-CuO-*IRF7*). The inducible vector co-expresses the repressor, CymR and puromycin resistance gene driven by consecutive EF1 α promoter (Figure 1A). Consistently expressed CymR binds to the CuO promoter to repress gene expression and addition of cumate changes the conformation of CymR which then turns on the gene switch by removing the repressor from the promoter.

Inducible IRF7 Overexpression Cell Line Establishment

Immortalized chicken embryonic fibroblast DF-1 cells (ATCC, Manassas, VA, United States) were cultured in Dulbecco's modified Eagle's medium (Thermo Fisher Scientific, Waltham, MA, United States) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA, United States), 1x Antibiotic-antimycotic (Thermo Fisher Scientific, Waltham, MA, United States), and incubated at 37°C in a humidified atmosphere containing 5% CO₂. Empty vector (Control) or chicken IRF7 inducible expression vector (pB-CuO-*IRF7*) was co-transfected with pB transposase plasmid into DF-1 cell lines using the Lipofectamine 3000 reagent (Thermo Fisher Scientific, Waltham, MA, United States) according to the manufacturer's protocol for efficient integration. Puromycin (3 μ g/ml) was added to the culture media 48 h after transfection and stably integrated cell lines were selected for 2 weeks. To induce *IRF7* expression, cumate (4-Isopropylbenzoic acid, Sigma-Aldrich, St. Louis, MO, United States) was added to the culture media at 12 h after seeding for 24 h followed by subsequent *in vitro* experiments.

Quantitative Reverse Transcriptase PCR

Total RNA was isolated from approximately 1 million cells using Direct-zol RNA MiniPrep Kit (Zymo Research, Irvine, CA, United States) and complement DNA (cDNA) was synthesized from total RNA (500 ng) using Verso cDNA Synthesis Kit (Thermo Fisher Scientific, Waltham, MA, United States). Quantitative reverse transcriptase PCR (qRT-PCR) was performed using the Applied Biosystems 7500 Fast Real-Time PCR System (Life Technologies, Grand Island, NY, United States) with SYBR Select Master Mix (Life Technologies, Grand Island, NY, United States). *IRF7*, *IFNA*, *IFNB* expression was normalized to the chicken glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene using the $\Delta\Delta C_T$ method (Livak and Schmittgen, 2001; Kim and Zhou, 2015).

Virus and *in vitro* AIV Infection

A/Chicken/California/2000 (H6N2) and A/Chicken/California/1999 (H10N7) low pathogenic avian influenza virus (LPAIV) strains were kindly provided by Dr. Rodrigo Gallardo (University of California, Davis, CA, United States) and Dr. Peter Woolcock [University of California, Davis, California Animal Health and Food Safety (CAHFS)], respectively. Each LPAIV was propagated in Madin-Darby Canine Kidney (MDCK) cells as described in Einfeld et al. (2014). All *in vitro* AIV infections were performed using CellBIND 12 well tissue culture plate (Corning, NY, United States) with 1×10^6 cells per well at the seeding. For viral replication kinetics phenotype, established DF-1 cell lines were induced and infected with either

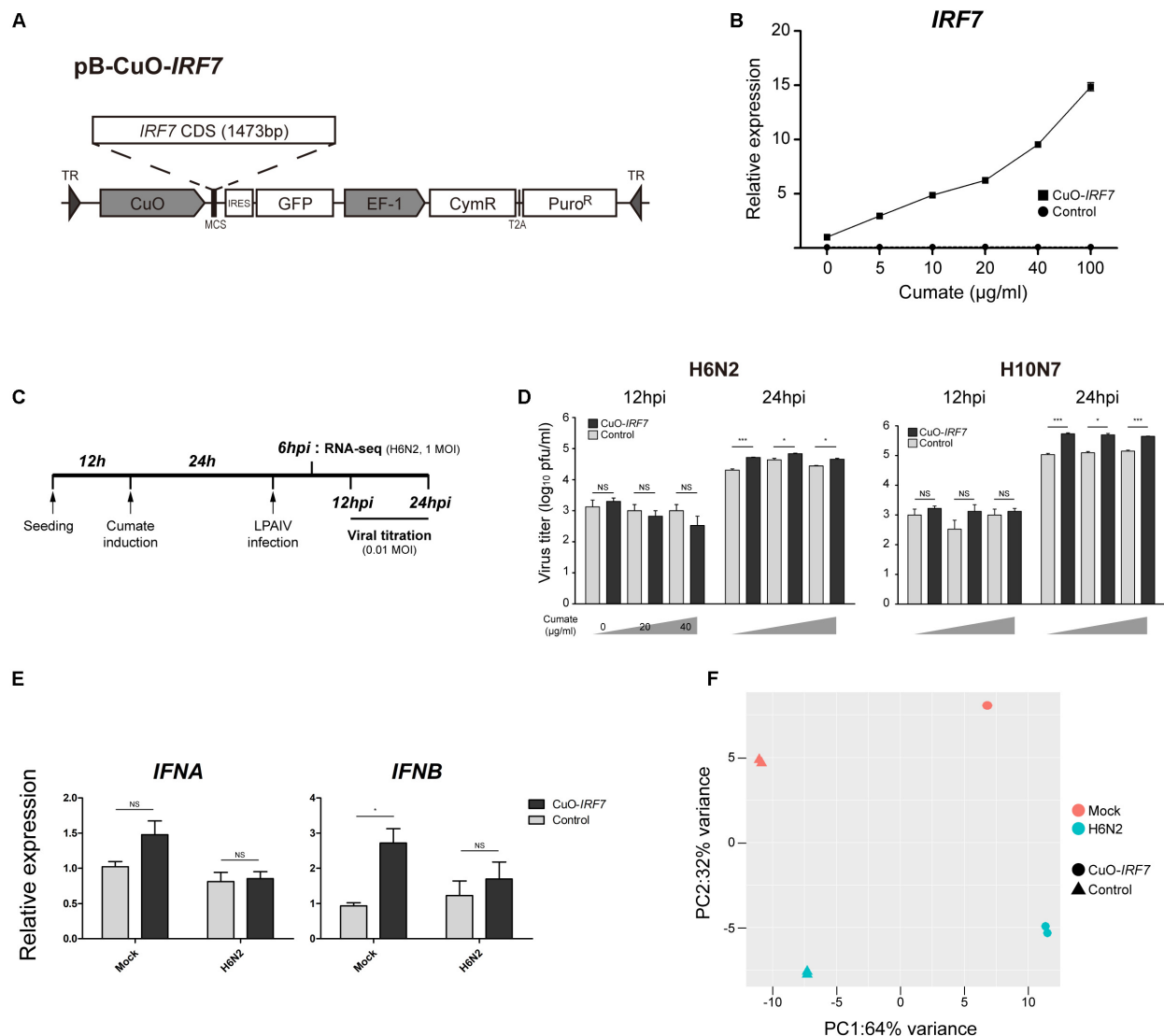


FIGURE 1 | Characterization of IRF7 inducible overexpression in DF-1 cell line **(A)** Schematic diagram of IRF7 inducible overexpression plasmid (CuO-IRF7). Empty vector plasmid was used to generate proper control (Control). **(B)** Titratable IRF7 expression induction. IRF7 expression induction was measured by qRT-PCR at 24 h post-induction of cumate. All data are shown as mean \pm SEM from three biological replicates. **(C)** Experimental timeline of *in vitro* low pathogenic avian influenza virus (LPAIV) infection. CuO-IRF7 and Control cell lines were induced by various level of cumate (0, 20, 40 µg/ml) at 12 h after passing for 24 h. Induced cell lines were infected with H6N2 or H10N7 [0.01 MOI (multiplicity of infection)] and progeny viral titer was measured by plaque assay at 12 and 24 hpi. For RNA-seq, cell lines were infected by either mock or H6N2 with an MOI of 1 and cells were harvested at 6 hpi. **(D)** Progeny viral production in the media was measured by plaque assay. All data are shown as mean \pm SEM ($N = 3$, * $p < 0.05$, *** $p < 0.001$, NS: not significant; Two tailed T -test). **(E)** Expression of type I IFNs in the CuO-IRF7 and control cell lines in mock and upon H6N2 infection (6 hpi, 1 MOI). Relative expression levels of IFNA and IFNB were measured by qRT-PCR. All data are shown as mean \pm SEM from three biological replicates (* $p < 0.05$, NS, not significant; Two tailed T -test). **(F)** Principal component analysis (PCA) of cDNA libraries. Replicates from each experimental group is separated by IRF7 expression level and H6N2 infection.

H6N2 or H10N7 at multiplicity of infection (MOI) of 0.01 with 0.05 µg/ml TPCK-trypsin in DMEM. Culture supernatants were collected at 0, 12, and 24 h post-infection (hpi) and the viral titer of each sample was measured by the plaque assay using MDCK cells (Huprikar and Rabinowitz, 1980). For the transcriptome profiling study, non-induced cell lines were infected with either mock or H6N2 at an MOI of 1 with 0.05 µg/ml TPCK-trypsin in DMEM. Trizol reagent (Thermo Fisher Scientific, Waltham, MA, United States) was directly added to the cell monolayer after

washed twice with PBS to extract the total RNA at 6 hpi which the IRF7 expression level starts to peak upon AIV infection (Kim and Zhou, 2015).

RNA Sequencing and Data Analysis

A total of eight cDNA libraries were prepared from two biological replicates of each group (mock or H6N2 infected Control or CuO-IRF7). RNA sequencing libraries were prepared from poly-adenylated RNA and sequenced with Illumina HiSeq4000

which generated over 20 million 150 bp paired-end reads per sample. The read files from RNA-seq analysis have been deposited in NCBI's Gene Expression Omnibus with GEO Series accession number GSE115131. We checked the quality of each library by fastQC (version 0.11.6) and trimmed the adaptor sequence by TrimGalore (version 0.4.5). We aligned the trimmed fastq files to the galGal5 chicken genome using STAR aligner (version 2.6.0) with NCBI annotation release 103 (Dobin et al., 2013). Unmapped reads were aligned against H6N2 (A/chicken/CA/6643/2001) genome (Webby et al., 2002). Raw read counts were extracted by HTSeq (version 0.9.0) from each aligned bam files and used to identify differentially expressed genes (DEGs) (Anders et al., 2015). Both DESeq2 and EdgeR R packages were used to identify DEGs and the DEG sets from the both packages were combined (false discovery rate (FDR) < 0.1% in any one of the packages) (Robinson et al., 2010; McCarthy et al., 2012; Love et al., 2014). The combined DEG lists were further filtered by removing the low expression genes if both samples of given comparison had fragments per kilobase per million reads (FPKM) value less than 1. Functional annotations for significantly DEGs were performed using DAVID 6.8 (Dennis et al., 2003; Huang da et al., 2009). The enriched gene ontology (GO) terms on biological processes and the pathways obtained from DAVID functional analysis were filtered for significance by gene count ≥ 5 and p -value < 0.05.

Hoechst 33342/Propidium Iodide Staining

Each control or CuO-*IRF7* cells were seeded at 1×10^5 cells/well to attachment factor protein (Thermo Fisher Scientific, Waltham, MA, United States) treated μ -Plate 96 Well glass bottom plate (ibidi GmbH, Germany) before 24 h of infection and infected 1MOI with three independent replicates. At 6 and 12 hpi, each well was washed twice with PBS and stained with 1 μ g/ml each of Hoechst 33342 (Thermo Fisher Scientific, Waltham, MA, United States) and propidium iodide (Thermo Fisher Scientific, Waltham, MA, United States) in PBS at room temperature for 10 min. Two images were taken from each replicate by using 20 \times objective lens and captured by Nikon NIS 3.0 software. A total number of cells (blue) and dead cells (red) were counted using ImageJ (Wayne Rasband, National Institutes of Health, United States) software from all images.

RESULTS

Inducible Over-Expression of IRF7 in DF-1 Cell Line

To precisely control the expression level of IRF7 *in vitro*, stable IRF7 inducible overexpression (CuO-*IRF7*) DF-1 cells were established using a cumate inducible vector (Figure 1A). Empty vector control cells (Control) which have identical expression cassettes without IRF7 CDS were also generated to exclude potential random gene disruption noise due to vector integration. We induced established cell lines with various cumate concentrations from 0 to 100 μ g/ml for 24 h to titer

the expression level of IRF7 by qRT-PCR (Figure 1B). There was approximately 10-fold higher IRF7 expression in CuO-*IRF7* cells compared to the control cells without induction (Supplementary Figure S1). Titratable *IRF7* expression was observed in CuO-*IRF7* cells as cumate concentration increases (up to 15-fold upregulation vs. non-induced cells) and the induction levels of *IRF7* were highly reproducible. In contrast, IRF7 expression level was not affected by cumate induction in control cells.

LPAIV Virus Replication in the Induced Cell Lines

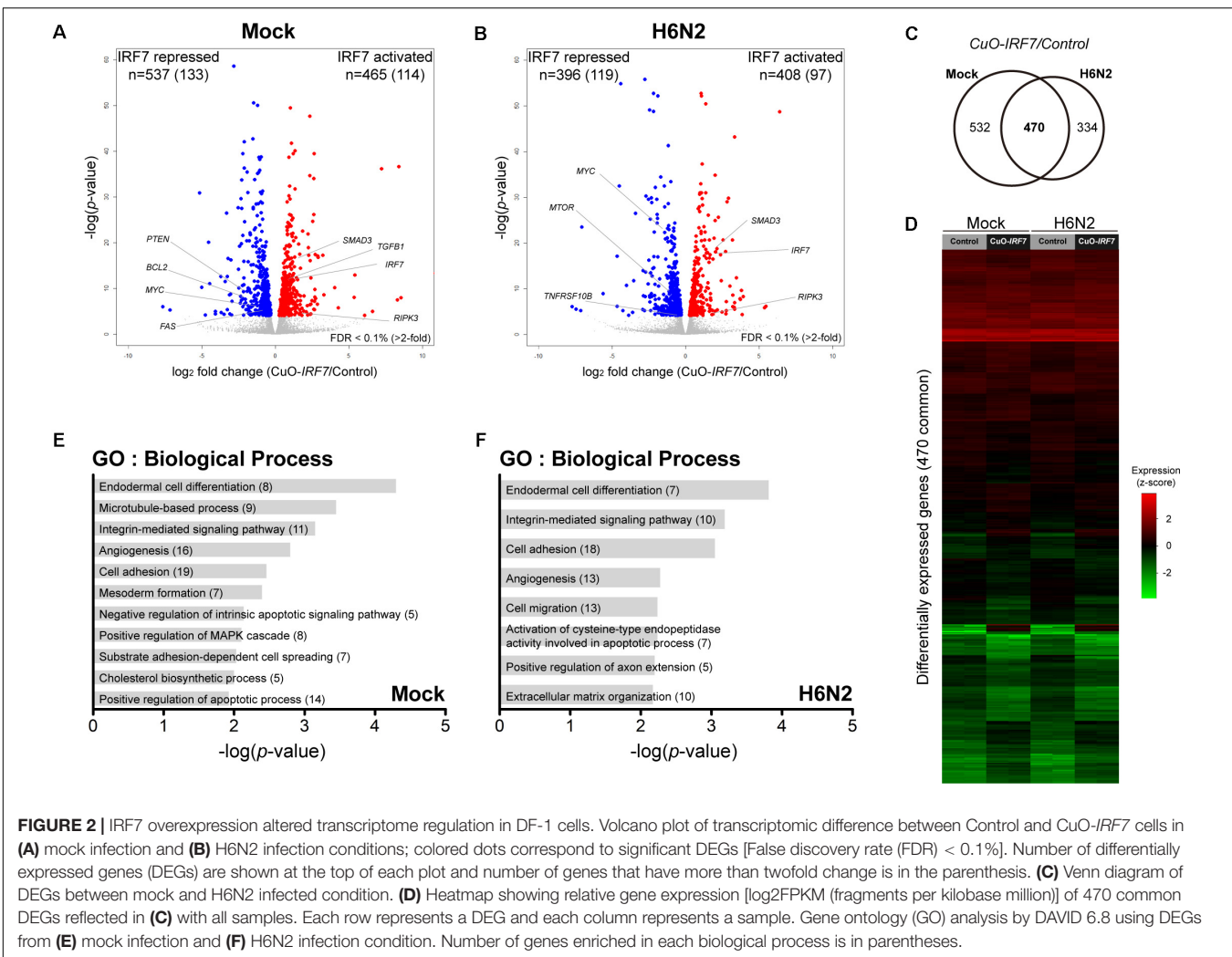
Two LPAI virus strains H6N2 and H10N7 at an MOI of 0.01 were used to infect the cumate induced cell lines, and the correlation between IRF7 expression level and the AIV replication was analyzed. Both CuO-*IRF7* and control cell lines were induced with 0, 20, or 40 μ g/ml of cumate for 24 h which correspond to approximately 10, 50, and 100-fold overexpression of *IRF7* in CuO-*IRF7* cells compared to the control cells at the time of infection. Then, AIV replication dynamics were analyzed at 12 and 24 h post-infection (hpi) by MDCK based plaque assay (Figure 1C). At 12 hpi, there was no significant difference in viral replication observed between *IRF7* overexpression and control cells for both LPAI virus strains at any level of induction. At 24 hpi, we observed significantly increased viral replication in *IRF7* overexpressed cell lines compared to the control cell lines from the both virus strains with a range of 1.6-fold (H6N2, 20 μ g/ml) up to fivefold (H10N7, 0 μ g/ml) (Figure 1D). However, we did not observe any significant correlation between the induced IRF7 expression level and the viral titer in the overexpression cells.

Type I IFN Regulation by IRF7

We additionally measured the expression levels of type I IFNs in mock or H6N2 infected (1 MOI) control and overexpression cells to examine the type I IFN regulator role of IRF7 (Figure 1E). In mock infected cell lines overexpression of IRF7 resulted almost threefold upregulation of IFNB whereas IFNA level did not show significant upregulation. There was no significant difference in expression of *IFNA* and *IFNB* between the Control and CuO-*IRF7* cells upon infection despite upregulation of IRF7 upon infection in each cell line (Control: 1.44-fold, FDR = 0.039; CuO-*IRF7*: 1.31-fold, FDR = 0.098, DESeq2).

Transcriptome Analysis

To identify the candidate genes and signaling pathways that IRF7 may regulate during AIV infection, we performed RNA-seq analysis on four different experimental groups of cells. There was no significant difference in viral replication among different IRF7 overexpression levels, thus non-induced DF-1 cells with either mock or H6N2 infection treatments were used and harvested at 6 hpi (Figure 1C). Each condition had 2 independent biological replicates and principal component analysis showed that the replicates in each condition grouped together by either IRF7 expression level or H6N2 infection condition (Figure 1F). The alignment rate against galGal5 reference genome was on



average of 84.98% from mock infection libraries and decreased to 71.57% with H6N2 infection as approximately 15% of the total reads were mapped to the viral genome in H6N2 infected cell lines.

Differential Expression Analysis of IRF7 Overexpression

First, by directly contrasting the transcriptomes between the control and IRF7 overexpression cell lines, we were able to identify the DEGs that were potentially regulated by IRF7 either at basal condition or upon H6N2 infection. There were 1,002 DEGs (465 up-, 537 down-regulated) in mock condition and 804 DEGs (408 up-, and 396 down-regulated) in H6N2 infection (Figures 2A,B and Supplementary Table S1). Comparison of the two DEG lists showed 470 genes were overlapped between mock and infection conditions (Figure 2C). The gene expression heatmap showed distinct gene expression pattern difference between control and CuO-IRF7 cells among the common 470 DEGs (Figure 2D). Gene ontology (GO) analysis from all DEGs revealed enriched functions involved

in cell structural integrity or cellular assembly such as cell adhesion, extracellular matrix (ECM) organization, integrin signaling mediated processes, and apoptosis (Figures 2E,F). Furthermore, we performed pathway analysis with the same gene sets to examine which pathways could be potentially altered as a result of IRF7 overexpression. Antibiotic and steroid biosynthesis pathways were significantly enriched in the mock condition while arginine, proline, and pyrimidine metabolisms were significantly enriched upon infection (Figures 3A,B). In addition, cell structure and ECM related pathways were enriched in both conditions (Figures 3A,B). Pathway analysis also revealed that FoxO signaling, TGF-beta signaling and PPAR signaling pathways, which are related to the immuno-regulation, could be possibly regulated by IRF7 (Figure 3A). DEGs in FoxO signaling and TGF-beta signaling pathways are presented in Supplementary Figure S2. Heatmaps with fold changes of all individual DEGs from significantly enriched pathways are presented in Figures 3C–H). Most of the integrin (ITG), laminin (LAM), and collagen (COL) DEGs were up-regulated in both conditions except for *ITGA9*, *LAMB1*, and *COL6A2*, which were down-regulated DEGs (Figure 3D). *TGFB1* and

SMAD3 were up-regulated whereas *PTEN* and *MYC* were down-regulated in the TGF-beta signaling pathway in both conditions (Figure 3E).

IRF7 Overexpression Increased Cell Death Upon AIV Infection

Based on the enriched GO terms and signaling pathways, we further investigated the effect of IRF7 overexpression on cell viability using Hoechst 33342/PI nucleic acid staining fluorescence microscopy. While Hoechst dyes can penetrate the living cells, propidium iodide only stains dead cells. Representative images of Hoechst 33342/PI staining are shown in Figure 4A (6 hpi) and Figure 4B (12 hpi). Approximately twice as many CuO-IRF7 cells ($48.4 \pm 5.3\%$) died upon H6N2 infection at 6dpi as control cells ($24.2 \pm 4.7\%$) at 6 hpi (Figure 4C).

Role of IRF7 in the Host Response Against H6N2 Infection

Next, we compared the host responses against H6N2 infection between the cell lines (Control/Mock vs. Control/H6N2 and CuO-IRF7/Mock vs. CuO-IRF7/ H6N2) to investigate which genes and pathways were potentially modulated by IRF7 upon infection (Supplementary Table S2). We identified 564 activated genes and 886 repressed genes in control cells upon H6N2 infection (Figure 5A) and a similar number of DEGs were identified from overexpression cells (Figure 5B; 557 activated genes, 735 repressed genes). There were 704 genes that were common between the two contrasts, and 746 and 588 genes were unique DEGs to the control and CuO-IRF7 contrast, respectively (Figure 5C). GO and pathway analysis using DEGs from the control contrast represented the molecular signature of host response against AIV infection (Figures 5D,E) while the CuO-IRF7 contrast could identify unique genes and signaling pathways associated with AIV infection that were potentially regulated by IRF7 (Figures 5F,G). Yet, despite the difference in functional annotation observed between the two contrasts, the individual genes did not show dramatic fold change differences between them (Supplementary Figure S3).

Then, we further analyzed the interaction of IRF7 overexpression by H6N2 infection to identify the genes that were differentially regulated during the infection by IRF7 (Supplementary Table S3). The top 50 interaction genes from a total of 350 DEGs (FDR < 5%) are listed in Figure 6A. GO analysis (Figure 6B) demonstrated that IRF7 overexpression could affect genes in cellular assembly, organization and structural functions as well as apoptosis. Pathway analysis (Figure 6C) also showed consistency of the candidate pathways enriched from previous analyses shown in Figures 3A,B, 5F,G. Of particular note, Janus kinase/signal transducers and activators of the transcription (JAK-STAT) signaling pathway were significantly represented in this analysis which was not enriched in any of the above direct contrast functional annotations (Figure 6D). In addition, the TGF-beta signaling pathway was also significantly enriched (Figure 6E).

DISCUSSION

Here, we employed a functional genomics approach to investigate the functional role of chicken IRF7 in the host innate immune response to AIV infection by generating an *in vitro* inducible overexpression model followed by whole transcriptome sequencing. Overexpression of IRF7 resulted in higher viral replication as well as greater cell death in our *in vitro* model. The transcriptome analysis suggested that chicken IRF7 might be involved in modulating a wide range of cellular processes including programmed cell death via the TGF-beta, FoxO, and JAK-STAT signaling pathways.

In this study, a cumate inducible system was applied to fine control the overexpression of IRF7 (a 10-fold change by qRT-PCR, and a twofold change by RNA-seq) and its range may better reflect actual physiological expression of IRF7 (Figure 1B) compared to the constitutive overexpression system we developed in our previous study that resulted in almost 200-fold overexpression (Kim and Zhou, 2015). Even with this substantially lower IRF7 overexpression level compared to the previous study, we were able to observe the significant *IFNB* upregulation (Figure 1E). This further supports a regulatory role of IRF7 on type I IFNs in chickens (Kim and Zhou, 2015). Furthermore, the relatively lower level of upregulation or repression patterns of IFNs upon H6N2 infection despite upregulation of IRF7 could suggest that manipulation of the host immune system by AIV affects the link between the IRF7 and type I IFNs (Alcami and Koszinowski, 2000).

Interestingly, IRF7 overexpression resulted in significantly higher viral titer in the overexpression cells than control cells regardless of IRF7 induction level (Figure 1D). However, it is yet to be determined whether the higher progeny viral titer resulting from IRF7 overexpression is detrimental or beneficial to the host as we also observed increased levels of induced cell death (Figure 4). Programmed cell death such as apoptosis and necroptosis are host defense strategies known to limit viral infection by eliminating the environment for viruses to replicate (Upton and Chan, 2014; Orzalli and Kagan, 2017), while influenza viruses adapted to inhibit the apoptosis process to benefit their survival through modulating the host response (Zhironov et al., 2002; Ehrhardt et al., 2007). On the other hand, influenza virus can induce immense host cell death for effective replication and transmission that results in morbidity, pathogenesis and virulence (Brydon et al., 2005; Ludwig et al., 2006). Further *in vivo* investigation could test the hypothesis that IRF7 overexpression was beneficial to the host by promoting the programmed cell death to limit the virus.

Pathway analysis suggested both the TGF-beta and FoxO signaling pathways as potential mechanisms that could be modulated by chicken IRF7 in the host response to AIV. The TGF-beta signaling pathway has diverse functions in cells and tissues, including cell-cycle control, differentiation, extracellular matrix formation, and apoptotic activation (Massague and Chen, 2000; Schuster and Kriegstein, 2002). The FoxO signaling pathway also plays important roles in metabolism, stress resistance, cellular proliferation, and apoptosis by transcription

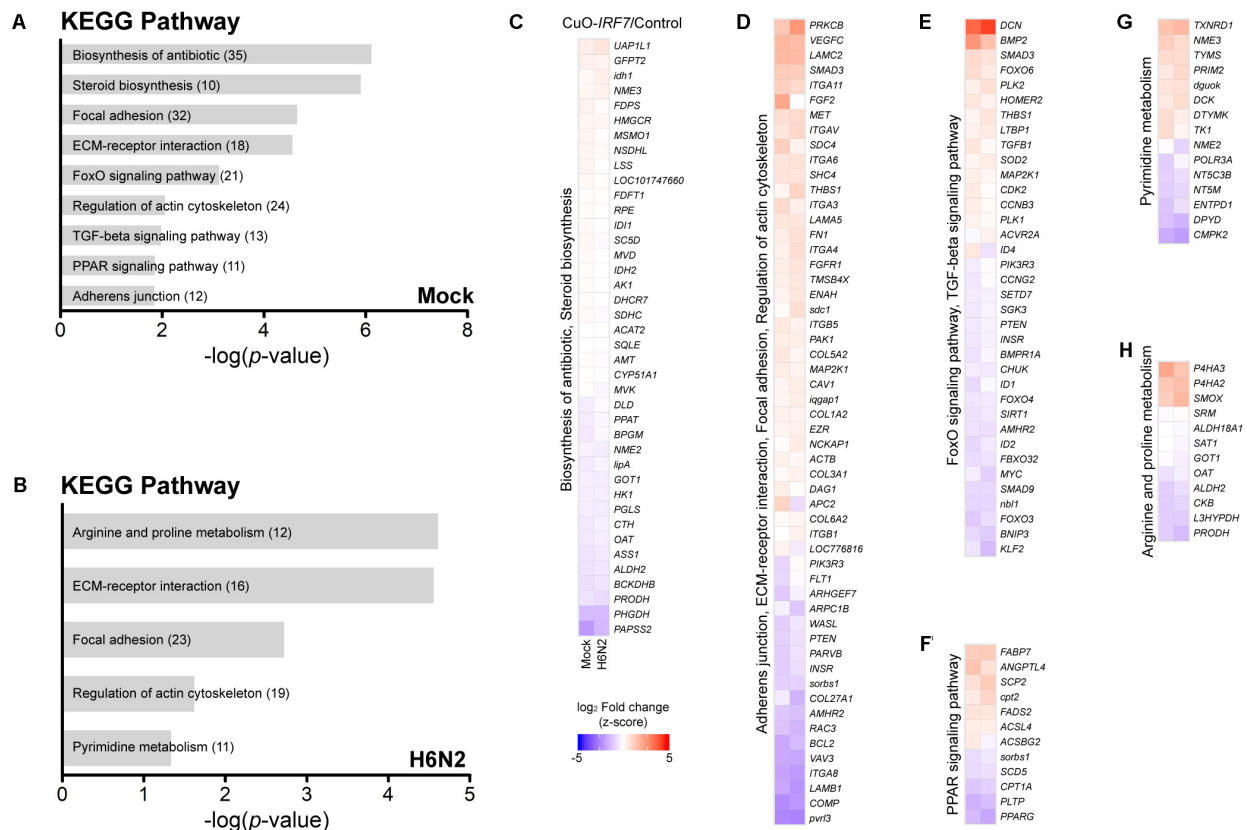


FIGURE 3 | Pathway analysis of DEGs and their expression patterns. Pathway analysis enrichment using DEGs between Control and CuO-IRF7 cells in (A) mock infection and (B) H6N2 infection conditions. Number of genes in each pathway is in parentheses. (C–H) Heatmaps showing the expression fold change of DEGs in significantly enriched pathways in mock infection (left column) and H6N2 infection (right column). Pathways were combined based on the common genes across the pathways. Lists of DEGs from both conditions in each pathway were combined to generate each heatmap.

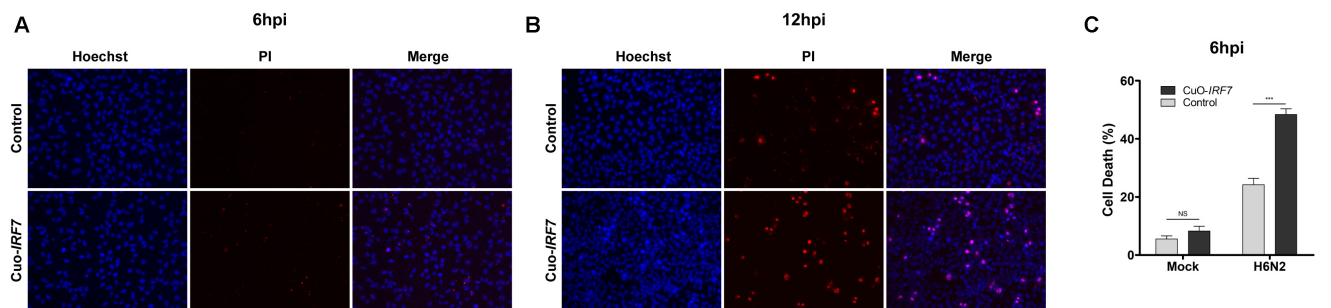
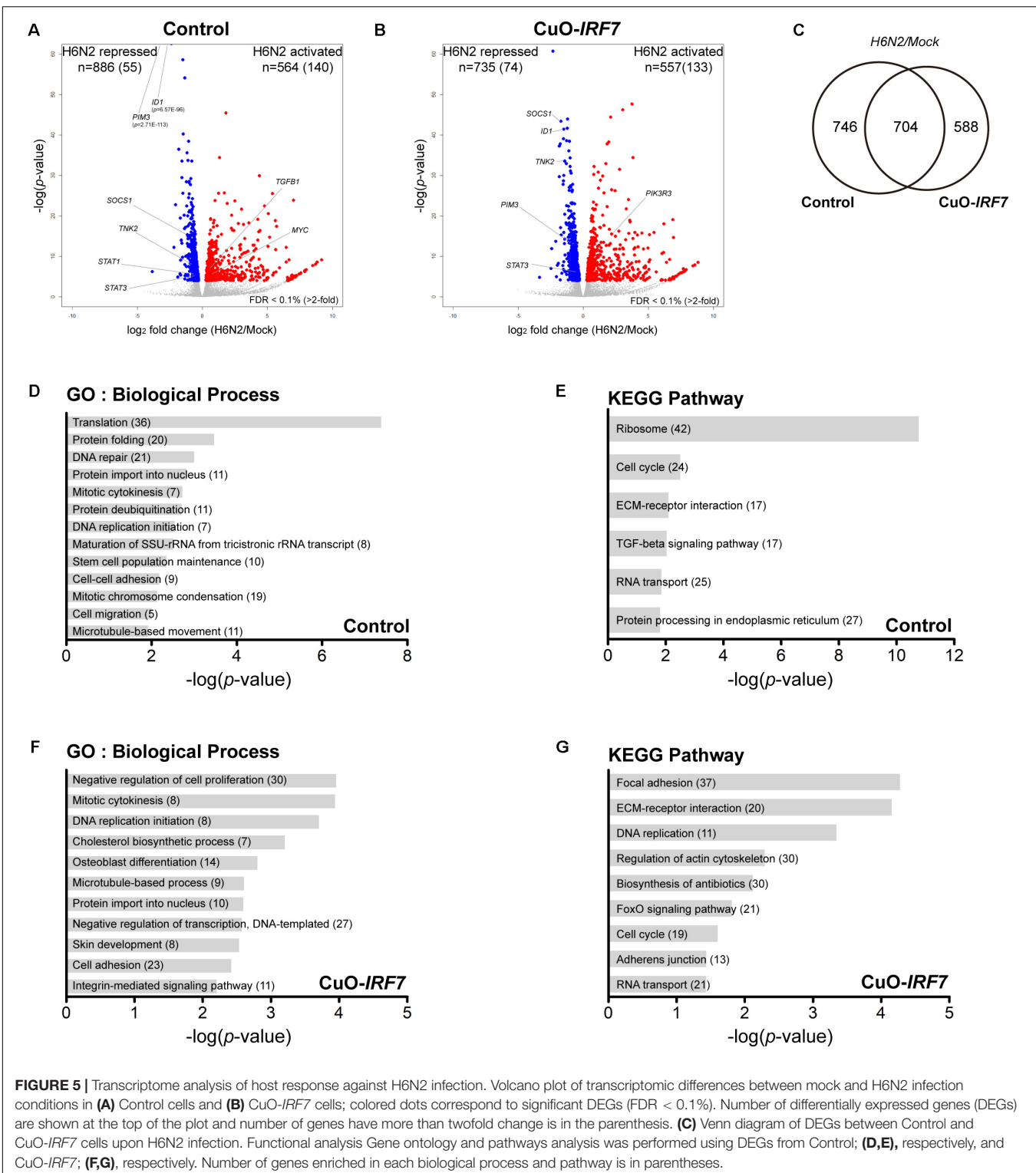


FIGURE 4 | IRF7 overexpression resulted more cell death. Representative images of Hoechst 33342 and Propidium Iodide (PI) double nuclear staining of control and CuO-IRF7 cell lines at (A) 6 h ($N = 6$) and (B) 12 h ($N = 3$) post-infection with H6N2. Images were visualized by 20x subjective lens. (C) The percentage of dead cells (PI positive; red) compared with total cells (Hoechst positive; blue) for Control and CuO-IRF7 cells with mock and H6N2 condition at 6 hpi. All data are shown as mean \pm SEM from six replicates. (***) $p < 0.001$, Two tailed T -test).

factor family Forkhead box (FOX) proteins. TGF-beta and FoxO signaling pathways are often considered together as their mechanisms of action are closely associated (Naka et al., 2010; Zhang et al., 2011; Wang Z. et al., 2014). TGF-beta induces apoptosis by SMAD-dependent manner (Jang et al., 2002; Schuster and Kriegstein, 2002), and we observed

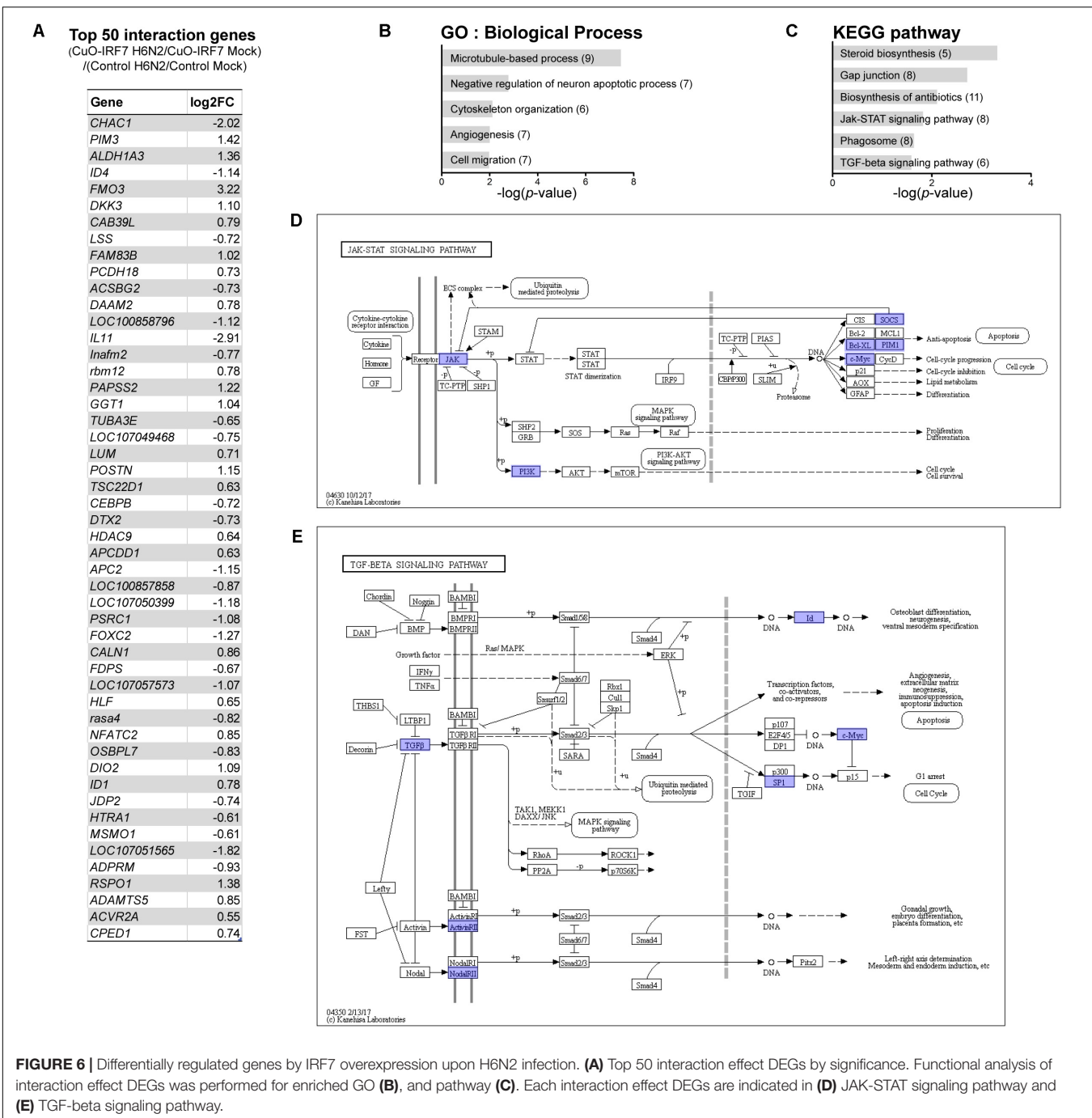
upregulation of *TGFB1* and *SMAD3* and downregulation of anti-apoptotic gene *BCL2* by IRF7 overexpression. This reflects the increased apoptosis and implies IRF7 regulation of the process (Figures 2A,B, 3E and Supplementary Figure S2). Qing et al. (2004) suggested the regulation of IRF7 function by TGF-beta/Smad3 signaling as a possible mechanism of the host type I



IFN response in the mouse embryonic fibroblasts, yet our results suggest a possible regulation of TGF-beta/Smad3 signaling by chicken IRF7.

FOXO3 was shown as a negative regulator of IRF7 gene transcription in the mouse macrophages in which

FOXO3 directly binds at the IRF7 promoter and could control its transcription (Litvak et al., 2012). Another study suggested a negative regulatory role of FOXO1 in the cellular antiviral response by promoting the ubiquitination of IRF3 and subsequent IRF3 protein degradation (Lei et al., 2013).



Both studies suggest the possible regulatory circuit that some FOXO proteins controlling IRF3 or IRF7 to prevent excessive innate immune response that could result pathological outcome. Differential expression of FOXO3, FOXO4 and FOXO6 as a result of the IRF7 overexpression (**Figure 3E** and **Supplementary Figure S2A**) may suggest possible conservation of the feedback circuits controlling the antiviral response associated with FOXO transcription factors in chickens.

The death receptor (DR) signaling pathway is another arm of the extrinsic apoptotic process (Orzalli and Kagan, 2017).

There have been reports regarding the detrimental effects of IFNs in influenza virus infection in which excessive IFN levels lead to severe damage on the host (McNab et al., 2015). Influenza virus susceptible mouse strains were found to have a stronger and more sustained type I IFN signal than resistant strains and antagonizing the type I IFN signal in susceptible strains improved host survival and reduced inflammation (Davidson et al., 2014). It was suggested that the disease-promoting effects of IFN are possibly mediated by upregulation of apoptosis-inducing proteins such as TNF-related apoptosis-inducing ligand

(TRAIL) and its receptor DR5 (TNFRSF10B) or Fas cell surface death receptor (FAS) which could lead to the tissue damage in somatic cells and immunosuppression on immune cells (Fujikura et al., 2013; Hogner et al., 2013; Davidson et al., 2014). In our study, *TRAIL* was not differentially expressed and downregulation of *DR5* and *FAS* (**Figures 2A,B**) were observed due to IRF7 overexpression, which suggests that the IRF7 mediated apoptosis may not utilize the DEATH receptor mediated mechanism.

Necroptosis is now recognized as an alternative to apoptosis as a mechanism of controlled cell death. Necroptosis has a distinct regulation mechanism compared to unintentional cell death by necrosis and it is a better inducer of a strong proinflammatory response that is crucial to the host immune response when compared to the apoptosis (Mocarski et al., 2015; Orzalli and Kagan, 2017). A protective role of Receptor Interacting Serine/Threonine Kinase 3 (RIPK3) has been reported against viral infections including influenza A virus and RIPK3 is known as a key factor upon viral infection that determines whether the infected cells undergo apoptosis or necroptosis (Newton et al., 2014; Nogusa et al., 2016; Thapa et al., 2016; Daniels et al., 2017). In this study, *RIPK3* was also up-regulated by IRF7 overexpression and along with the increased cell death, this might suggest a novel functional role of chicken IRF7 as a potential regulator in the necroptosis pathway against AIV infection.

In addition, chicken IRF7 may regulate genes involved in the cell structural integrity or cellular assembly such as cell adhesion, ECM organization, and adherens junctions (**Figures 2E,F, 3A,B**) which have a wide range of functions in the host response to viral infection. Actin cytoskeleton plays an important role in the entry of influenza virus into cells and proper assembly of viral particles (Sun and Whittaker, 2013; Kumakura et al., 2015). ECMs are also critical across many stages of the viral life cycle, including viral entry, transmission, and exit (Stavolone and Lionetti, 2017). Integrin mediated cell adhesion to ECM is essential for survival of many cell types (Meredith and Schwartz, 1997), and apoptotic cells undergo distinct morphological changes characterized by cell and nucleus shrinkage as well as disassembly into apoptotic bodies which are associated with structural proteins such as actins and laminins (Saraste and Pulkki, 2000; Suzanne and Steller, 2009). Actin initiates and mediates mammalian apoptosis via the intrinsic and extrinsic pathways and final degradation of actin filaments amplifies the apoptosis signaling cascade. Differential regulation of these structural proteins (mostly up-regulated by IRF7 overexpression, **Figure 3D**) may contribute to the increased apoptosis and viral production yet the precise mechanism remains to be further elucidated.

Furthermore, interaction analysis of IRF7 overexpression and H6N2 infection was performed to discover the genes that were differentially regulated during the infection due to IRF7 overexpression. The analysis not only reinforced our findings from the direct contrast analyses (**Figures 2, 3**), which suggested the modulator role of IRF7 in a wide range of host responses to H6N2 infection, but also discovered additional genes and pathways that were possibly modulated by IRF7. For example, the JAK-STAT signaling pathway was not enriched in any of direct contrast analyses but was significantly enriched from the

interaction DEGs. The JAK-STAT pathway is one of the key pathways in the type I IFN response and employs interferon-stimulated genes to inhibit virus infection by targeting the viral life cycle and regulate the host processes (Schneider et al., 2014; Majoros et al., 2017). In our study, Tyrosine kinase 2 (TYK2, one of the JAKs) and its negative regulators SOCS1, SOCS3 as well as anti-apoptosis genes BCL-XL and PIM1 were differentially regulated by IRF7 during the H6N2 infection.

In sum, overexpression of IRF7 resulted in higher viral replication as well as increased cell death in DF-1 cell lines. Although it is unclear in our *in vitro* model if increased viral replication due to IRF7 overexpression is beneficial to the host or to the virus, our results suggest potential modulator function of chicken IRF7 in the programmed cell death via TGF-beta-FOXO signaling axis in the host response. In addition, we revealed sets of candidate genes that IRF7 might regulate in the cellular structure organization, highlighting cell-cell adhesion processes that play an important role in both host response and viral life cycle. Chicken IRF7 was also involved in metabolic pathways that have known functions in antiviral response. An ongoing complete loss-of-function study of IRF7 by CRISPR-Cas9 will expand our knowledge on potential regulatory role of IRF7-dependent and -independent pathways on AIV infection in poultry. A genome-wide IRF7 binding study by ChIP-seq would also help to expand our knowledge by determining whether the DEGs were directly modulated by IRF7 binding or by the cascade of downstream type I IFN responses.

AUTHOR CONTRIBUTIONS

TK and HZ conceived and designed the experiments, contributed reagents, materials, analysis tools, analyzed the data, and wrote the paper. TK performed the experiments.

FUNDING

This work was supported by United States Department of Agriculture (USDA), National Institute of Food and Agriculture (NIFA), Multistate Research Project NRSP8 and NC1170 (HZ), and the California Agricultural Experimental Station (HZ). TK was partially supported by Fellowships Grant Program Award #2017-67011-26762 from the USDA NIFA.

ACKNOWLEDGMENTS

Two LPAIV strains used in this study were gifts from the labs of Dr. Rodrigo Gallardo and Dr. Peter Woolcock. Ms. Ganrea Chanthavixay and Ms. Karen Tracy provided great inputs on revision of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00415/full#supplementary-material>

REFERENCES

- Alcami, A., and Koszinowski, U. H. (2000). Viral mechanisms of immune evasion. *Trends Microbiol.* 8, 410–418. doi: 10.1016/S0966-842X(00)01830-8
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Brydon, E. W., Morris, S. J., and Sweet, C. (2005). Role of apoptosis and cytokines in influenza virus morbidity. *FEMS Microbiol. Rev.* 29, 837–850. doi: 10.1016/j.femsre.2004.12.003
- Daniels, B. P., Snyder, A. G., Olsen, T. M., Orozco, S., Oguin, T. H. III, Tait, S. W. G., et al. (2017). RIPK3 restricts viral pathogenesis via cell death-independent neuroinflammation. *Cell* 169, 301.e11–313.e11. doi: 10.1016/j.cell.2017.03.011
- Davidson, S., Crotta, S., McCabe, T. M., and Wack, A. (2014). Pathogenic potential of interferon alphabeta in acute influenza infection. *Nat. Commun.* 5:3864. doi: 10.1038/ncomms4864
- de Weerd, N. A., Samarajiva, S. A., and Hertzog, P. J. (2007). Type I interferon receptors: biochemistry and biological functions. *J. Biol. Chem.* 282, 20053–20057. doi: 10.1074/jbc.R700006200
- Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:3. doi: 10.1186/gb-2003-4-5-p3
- Der, S. D., Zhou, A., Williams, B. R., and Silverman, R. H. (1998). Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15623–15628. doi: 10.1073/pnas.95.26.15623
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Downing, T., Cormican, P., O'Farrelly, C., Bradley, D. G., and Lloyd, A. T. (2009). Evidence of the adaptive evolution of immune genes in chicken. *BMC Res. Notes* 2:254. doi: 10.1186/1756-0500-2-254
- Ehrhardt, C., Wolff, T., Pleschka, S., Planz, O., Beermann, W., Bode, J. G., et al. (2007). Influenza A virus NS1 protein activates the PI3K/Akt pathway to mediate antiapoptotic signaling responses. *J. Virol.* 81, 3058–3067. doi: 10.1128/JVI.02082-06
- Eisfeld, A. J., Neumann, G., and Kawaoka, Y. (2014). Influenza A virus isolation, culture and identification. *Nat. Protoc.* 9, 2663–2681. doi: 10.1038/nprot.2014.180
- Fujikura, D., Chiba, S., Muramatsu, D., Kazumata, M., Nakayama, Y., Kawai, T., et al. (2013). Type-I interferon is critical for FasL expression on lung cells to determine the severity of influenza. *PLoS One* 8:e55321. doi: 10.1371/journal.pone.0055321
- Goossens, K. E., Ward, A. C., Lowenthal, J. W., and Bean, A. G. (2013). Chicken interferons, their receptors and interferon-stimulated genes. *Dev. Comp. Immunol.* 41, 370–376. doi: 10.1016/j.dci.2013.05.020
- Hogner, K., Wolff, T., Pleschka, S., Plog, S., Gruber, A. D., Kalinke, U., et al. (2013). Macrophage-expressed IFN-beta contributes to apoptotic alveolar epithelial cell injury in severe influenza virus pneumonia. *PLoS Pathog.* 9:e1003188. doi: 10.1371/journal.ppat.1003188
- Honda, K., and Taniguchi, T. (2006). IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.* 6, 644–658. doi: 10.1038/nri1900
- Honda, K., Yanai, H., Negishi, H., Asagiri, M., Sato, M., Mizutani, T., et al. (2005). IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* 434, 772–777. doi: 10.1038/nature03464
- Huang, B., Qi, Z. T., Xu, Z., and Nie, P. (2010). Global characterization of interferon regulatory factor (IRF) genes in vertebrates: glimpse of the diversification in evolution. *BMC Immunol.* 11:22. doi: 10.1186/1471-2172-11-22
- Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Huprikar, J., and Rabinowitz, S. (1980). A simplified plaque assay for influenza viruses in Madin-Darby kidney (MDCK) cells. *J. Virol. Methods* 1, 117–120. doi: 10.1016/0166-0934(80)90020-8
- Jang, C. W., Chen, C. H., Chen, C. C., Chen, J. Y., Su, Y. H., and Chen, R. H. (2002). TGF-beta induces apoptosis through Smad-mediated expression of DAP-kinase. *Nat. Cell Biol.* 4, 51–58. doi: 10.1038/ncb731
- Jhung, M. A., and Nelson, D. I. C. (2015). Centers for disease, and prevention, outbreaks of avian influenza A (H5N2), (H5N8), and (H5N1) among birds—United States, December 2014–January 2015. *MMWR Morb. Mortal. Wkly. Rep.* 64:111.
- Kim, T. H., and Zhou, H. (2015). Functional Analysis of chicken IRF7 in response to dsRNA analog poly(I:C) by integrating overexpression and knockdown. *PLoS One* 10:e0133450. doi: 10.1371/journal.pone.0133450
- Kumakura, M., Kawaguchi, A., and Nagata, K. (2015). Actin-myosin network is required for proper assembly of influenza virus particles. *Virology* 476, 141–150. doi: 10.1016/j.virol.2014.12.016
- Lei, C. Q., Zhang, Y., Xia, T., Jiang, L. Q., Zhong, B., and Shu, H. B. (2013). FoxO1 negatively regulates cellular antiviral response by promoting degradation of IRF3. *J. Biol. Chem.* 288, 12596–12604. doi: 10.1074/jbc.M112.444794
- Litvak, V., Ratushny, A. V., Lampano, A. E., Schmitz, F., Huang, A. C., Raman, A., et al. (2012). A FOXO3-IRF7 gene regulatory circuit limits inflammatory sequelae of antiviral responses. *Nature* 490, 421–425. doi: 10.1038/nature11428
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Ludwig, S., Pleschka, S., Planz, O., and Wolff, T. (2006). Ringing the alarm bells: signalling and apoptosis in influenza virus infected cells. *Cell Microbiol.* 8, 375–386. doi: 10.1111/j.1462-5822.2005.00678.x
- Magor, K. E., Miranzo Navarro, D., Barber, M. R., Petkau, K., Fleming-Canepa, X., Blyth, G. A., et al. (2013). Defense genes missing from the flight division. *Dev. Comp. Immunol.* 41, 377–388. doi: 10.1016/j.dci.2013.04.010
- Majoros, A., Platanitis, E., Kernbauer-Holz, E., Rosebrock, F., Muller, M., and Decker, T. (2017). Canonical and non-canonical aspects of JAK-STAT signaling: lessons from interferons for cytokine responses. *Front. Immunol.* 8:29. doi: 10.3389/fimmu.2017.00029
- Massague, J., and Chen, Y. G. (2000). Controlling TGF-beta signaling. *Genes Dev.* 14, 627–644.
- McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297. doi: 10.1093/nar/gks042
- McNab, F., Mayer-Barber, K., Sher, A., Wack, A., and O'Garra, A. (2015). Type I interferons in infectious disease. *Nat. Rev. Immunol.* 15, 87–103. doi: 10.1038/nri3787
- Meredith, J. E. Jr., and Schwartz, M. A. (1997). Integrins, adhesion and apoptosis. *Trends Cell Biol.* 7, 146–150. doi: 10.1016/S0966-8924(97)01002-7
- Mocarski, E. S., Guo, H., and Kaiser, W. J. (2015). Necroptosis: the trojan horse in cell autonomous antiviral host defense. *Virology* 479–480, 160–166. doi: 10.1016/j.virol.2015.03.016
- Naka, K., Hoshii, T., Muraguchi, T., Tadokoro, Y., Ooshio, T., Kondo, Y., et al. (2010). TGF-beta-FOXO signalling maintains leukaemia-initiating cells in chronic myeloid leukaemia. *Nature* 463, 676–680. doi: 10.1038/nature08734
- Newton, K., Dugger, D. L., Wickliffe, K. E., Kapoor, N., de Almagro, M. C., Vucic, D., et al. (2014). Activity of protein kinase RIPK3 determines whether cells die by necroptosis or apoptosis. *Science* 343, 1357–1360. doi: 10.1126/science.1249361
- Nogusa, S., Thapa, R. J., Dillon, C. P., Liedmann, S., Oguin, T. H. III, Ingram, J. P., et al. (2016). RIPK3 activates parallel pathways of MLKL-driven necroptosis and FADD-mediated apoptosis to protect against influenza A virus. *Cell Host Microbe* 20, 13–24. doi: 10.1016/j.chom.2016.05.011
- Olsen, B., Munster, V. J., Wallensten, A., Waldenstrom, J., Osterhaus, A. D., and Fouchier, R. A. (2006). Global patterns of influenza A virus in wild birds. *Science* 312, 384–388. doi: 10.1126/science.1122438
- Orzalli, M. H., and Kagan, J. C. (2017). Apoptosis and necroptosis as host defense strategies to prevent viral infection. *Trends Cell Biol.* 27, 800–809. doi: 10.1016/j.tcb.2017.05.007
- Qing, J., Liu, C., Choy, L., Wu, R. Y., Pagano, J. S., and Derynck, R. (2004). Transforming growth factor beta/Smad3 signaling regulates IRF-7 function and

- transcriptional activation of the beta interferon promoter. *Mol. Cell. Biol.* 24, 1411–1425. doi: 10.1128/MCB.24.3.1411-1425.2004
- Ramos, S., MacLachlan, M., and Melton, A. (2017). *Impacts of the 2014–2015 Highly Pathogenic Avian Influenza Outbreak on the Poultry, U. S. Sector*. Washington, D.C: Economic Research Service
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., et al. (2013). Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* 41, D1040–D1046. doi: 10.1093/nar/gks1215
- Santhakumar, D., Rubbenstroth, D., Martinez-Sobrido, L., and Munir, M. (2017). Avian interferons and their antiviral effectors. *Front. Immunol.* 8:49. doi: 10.3389/fimmu.2017.00049
- Saraste, A., and Pulkki, K. (2000). Morphologic and biochemical hallmarks of apoptosis. *Cardiovasc. Res.* 45, 528–537. doi: 10.1016/S0008-6363(99)00384-3
- Schneider, W. M., Chevillotte, M. D., and Rice, C. M. (2014). Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* 32, 513–545. doi: 10.1146/annurev-immunol-032713-120231
- Schuster, N., and Kriegelstein, K. (2002). Mechanisms of TGF-beta-mediated apoptosis. *Cell Tissue Res.* 307, 1–14. doi: 10.1007/s00441-001-0479-6
- Stavolone, L., and Lionetti, V. (2017). Extracellular matrix in plants and animals: hooks and locks for viruses. *Front. Microbiol.* 8:1760. doi: 10.3389/fmicb.2017.01760
- Sun, X., and Whittaker, G. R. (2013). Entry of influenza virus. *Adv. Exp. Med. Biol.* 790, 72–82. doi: 10.1007/978-1-4614-7651-1_4
- Suzanne, M., and Steller, H. (2009). Letting go: modification of cell adhesion during apoptosis. *J. Biol.* 8:49. doi: 10.1186/jbiol152
- Swayne, D. E. (2012). Impact of vaccines and vaccination on global control of avian influenza. *Avian Dis.* 56, 818–828. doi: 10.1637/10183-041012-Review.1
- Thapa, R. J., Ingram, J. P., Ragan, K. B., Nogusa, S., Boyd, D. F., Benitez, A. A., et al. (2016). DAI senses influenza a virus genomic RNA and activates RIPK3-dependent cell death. *Cell Host Microbe* 20, 674–681. doi: 10.1016/j.chom.2016.09.014
- Upton, J. W., and Chan, F. K. (2014). Staying alive: cell death in antiviral immunity. *Mol. Cell.* 54, 273–280. doi: 10.1016/j.molcel.2014.01.027
- Wang, Y., Zhou, Y., and Graves, D. T. (2014). FOXO transcription factors: their clinical significance and regulation. *Biomed. Res. Int.* 2014:925350. doi: 10.1155/2014/925350
- Wang, Z., Ning, Z., Sun, M., Gao, S., Kang, Y., Xie, P., et al. (2014). Interferon regulatory factor 7- (IRF7-) mediated immune response affects newcastle disease virus replication in chicken embryo fibroblasts. *Acta Vet. Hung.* 62, 500–511. doi: 10.1556/AVet.2014.023
- Webby, R. J., Woolcock, P. R., Krauss, S. L., and Webster, R. G. (2002). Reassortment and interspecies transmission of North American H6N2 influenza viruses. *Virology* 295, 44–53. doi: 10.1006/viro.2001.1341
- Zhang, X., Tang, N., Hadden, T. J., and Rishi, A. K. (2011). Akt, FoxO and regulation of apoptosis. *Biochim. Biophys. Acta* 1813, 1978–1986. doi: 10.1016/j.bbamcr.2011.03.010
- Zhirnov, O. P., Konakova, T. E., Wolff, T., and Klenk, H. D. (2002). NS1 protein of influenza a virus down-regulates apoptosis. *J. Virol.* 76, 1617–1625. doi: 10.1128/JVI.76.4.1617-1625.2002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kim and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Applications of Gene Editing in Chickens: A New Era Is on the Horizon

Hicham Sid and Benjamin Schusser*

Department of Animal Sciences, Reproductive Biotechnology, School of Life Sciences Weihenstephan, Technical University Munich, Freising, Germany

OPEN ACCESS

Edited by:

Mark S. Fife,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Robert Etches,
Crystal Bioscience, United States
Jiuzhou Song,
University of Maryland, College Park,
United States

*Correspondence:

Benjamin Schusser
benjamin.schusser@tum.de

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 29 May 2018

Accepted: 18 September 2018

Published: 09 October 2018

Citation:

Sid H and Schusser B (2018)
Applications of Gene Editing in
Chickens: A New Era Is on the
Horizon. *Front. Genet.* 9:456.
doi: 10.3389/fgene.2018.00456

The chicken represents a valuable model for research in the area of immunology, infectious diseases as well as developmental biology. Although it was the first livestock species to have its genome sequenced, there was no reverse genetic technology available to help understanding specific gene functions. Recently, homologous recombination was used to knockout the chicken immunoglobulin genes. Subsequent studies using immunoglobulin knockout birds helped to understand different aspects related to B cell development and antibody production. Furthermore, the latest advances in the field of genome editing including the CRISPR/Cas9 system allowed the introduction of site specific gene modifications in various animal species. Thus, it may provide a powerful tool for the generation of genetically modified chickens carrying resistance for certain pathogens. This was previously demonstrated by targeting the Trp38 region which was shown to be effective in the control of avian leukosis virus in chicken DF-1 cells. Herein we review the current and future prospects of gene editing and how it possibly contributes to the development of resistant chickens against infectious diseases.

Keywords: chicken, CRISPR/Cas9, transgenic, knockout, Diseases, Immunoglobulins

INTRODUCTION

The chicken represents an important source of protein worldwide and a valuable model for the study of developmental biology in vertebrates (Yasugi and Nakamura, 2000; Speedy, 2003). Chickens are constantly exposed to a plethora of pathogens threatening animal welfare as well as human health (Perdue and Swayne, 2005; Humphrey, 2006). Viral pathogens such as influenza A viruses can be transmitted to humans leading to death (Gao et al., 2013). Furthermore, bacterial agents such as *Campylobacter jejuni* and *Salmonella enteritidis* cause food borne illnesses in humans associated with digestive symptoms (Bryan and Doyle, 1995). More recently, using genetically modified chickens as a model for various research areas like developmental biology, immunology, physiology and neurology is gaining importance in the avian research community (Mozdziak and Petite, 2004; Stern, 2004, 2005). In addition, there is an increasing interest to generate genetically modified chickens resistant to specific pathogens, benefiting from the availability of gene manipulation techniques. This review focuses on the advances made in gene editing in chickens and the future perspectives including the generation of specific-pathogen-resistant birds.

STATE OF THE ART

Genetically modified animals have significantly contributed to our understanding of different aspects related to immunity, infectious diseases, neurology, behavior, and developmental biology (Yeh et al., 2002; Lyall et al., 2011; Lalonde et al., 2012; Pinkert, 2014; Park et al., 2017b). While mice were the first animals to be genetically modified (Costantini and Lacy, 1981; Gordon and Ruddle, 1981), pronuclear DNA microinjections allowed the introduction of foreign DNA leading to genetic modifications in livestock including rabbits, sheep and pigs (Hammer et al., 1985). Although this method was used for a long time, it did not allow the induction of targeted gene modifications and had the disadvantage of generating random integrations (Perleberg et al., 2018). The generation of knockout (KO) animals was achieved for the first time by gene targeting in embryonic stem cells (ES) (Evans and Kaufman, 1981; Thomas and Capecchi, 1987). Though the induction of the KO was successful, it had the disadvantage of low efficiency (Thomas and Capecchi, 1987). Due to the absence of true ES lines from farm animals and no solid evidence of germline transmission (Talbot and Blomberg, 2008; Soto and Ross, 2016), stable transfection of sheep somatic cells with human factor IX and neomycin resistance followed by nuclear transfer was the alternative to express foreign DNA in livestock (Schnieke et al., 1997) and afterwards for gene targeting (McCreath et al., 2000). At this time, the generation of KO livestock animals was possible by combining somatic cell nuclear transfer (SNTC) and homologous recombination (Lai et al., 2002; Nottle et al., 2007). The laborious procedure of these methods and the low efficiency for generating targeted KO was improved by homologous recombination (Houdebine, 2002) along with different nucleases (Carlson et al., 2012). The transcription activator-like effector nucleases (TALENs) are composed of series of repeats fused to non-specific FokI-cleavage domains that induce double-stranded DNA breaks upon dimerization (Gaj et al., 2013). More recently the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system made the process of specific DNA-targeting easier by using single guide RNAs (sgRNAs) (Jinek et al., 2012; Ran et al., 2013; Hsu et al., 2014). CRISPR/Cas9 is an adaptive immune system found in bacteria and archaeal species and uses small-non coding RNAs to guide the Cas9 nuclease to target sites resulting in DNA double-break (Jinek et al., 2012).

In comparison to mammals, difficulties were always associated with the generation of genetically modified chickens due to the complex structure of the chicken zygote (Mozdziak and Petite, 2004) and the different organization of the chick embryo compared to mammals (Stern, 1990). Over the past 30 years, different research groups paved the way for the generation of genetically modified chickens. Efforts were focused on the stable genomic integration of transgenes and obtaining the highest efficiency of germline transmission. While Petite and colleagues described the transfer of stage X embryo cells that led to germline transmission, it was not possible to genetically modify these cells and to re-introduce them as germline competent cells into the chicken embryo (Petite et al., 1990). Although ES were shown to provide a valuable tool for the generation of transgenic

mice (Kanatsu-Shinohara et al., 2003), no evidence of germline transmission using chicken ES was reported. Transferred chicken ES cells only contributed to somatic tissue but not to the germline.

The first genetically modified chicken was generated by the insertion of retroviral foreign DNA delivered by avian leukosis virus that was successfully integrated to the germline (Salter et al., 1987). The retroviral vector was injected into the yolk sac near to the developing blastoderm. Since then, various viral vectors have been used to generate genetically modified chickens (Hughes et al., 1986; Bosselman et al., 1989; Salter and Crittenden, 1989; Harvey and Ivarie, 2003; Mozdziak et al., 2003). Drawbacks of viral vectors, such as the replication of deficient viral particles and risks of recombination with wild type viruses, were avoided by plasmid-DNA microinjection into the chicken zygote (Love et al., 1994). The microinjection was done in the germinal disk and led to the generation of transgenic chickens expressing neomycin resistance and a reporter gene *lacZ* (Love et al., 1994). A total of 5.5% of the generated chicks survived to sexual maturity and later on, one rooster gave 3.4% transmission to his offspring (Love et al., 1994). The germline transmission of integrated transgenes was improved with lentiviral vectors (McGrew et al., 2004). McGrew and colleagues showed the possibility of transduction with lentiviral vectors in G0 birds. Founder cockerels were injected with different plasmids carrying different reporter genes including *LacZ* and eGFP (McGrew et al., 2004). Lentiviral vectors were injected into the subgerminal cavity of newly laid eggs. Ten of the founder males transmitted 4–45% of the foreign DNA to their offspring (McGrew et al., 2004). Lentiviral vectors offered for the first time the possibility to generate genetically modified chickens with a decent germline transmission efficiency. Nevertheless, the size of the transgene was still limited and precise edits were not possible.

Furthermore, the *in ovo* injection of the avian retroviral vector RCAS (replication-competent avian sarcoma-leukosis virus with a splice acceptor) carrying enhanced fluorescent protein (eGFP) into unincubated (stage X) blastoderms resulted in stable and widespread expression of eGFP in the embryos. Even though the gonads showed eGFP expression PGCs were eGFP negative indicating viral silencing (Smith et al., 2009).

Like in mammals, chicken primordial germ cells (PGCs) are precursors of gametes and a key element for sperm and oocytes development. At the early hours of embryonic development, PGCs are found in the germinal crescent and migrate afterwards (50–55 h) to the gonads (Kim et al., 2010; Kang et al., 2015) in order to produce sperm and oocytes upon sexual maturity (Fujimoto et al., 1976). The migration of PGCs was found to be greatly influenced by the chemokine stromal cell-derived factor-1 (SDF-1/CXCL12) and its receptor C-X-C chemokine receptor type 4 (CXCR4) (Stebler et al., 2004; Lee et al., 2017c).

The ability to culture PGCs was a milestone in the process of generating transgenic chickens. Genetic modification of PGCs and their subsequent reintroduction into the embryonic vasculature resolved many issues and problems observed with previously established methods. Van de Lavoie and colleagues used BRL or STO feeder cells to cultivate PGCs for up to 217 days. PGCs were shown to retain the germline characteristics by analyzing various germline markers including the chicken vasa

homolog (CVH) and were cryoconserved using conventional techniques (Van De Lavoie et al., 2006). PGC-culture was optimized afterwards by Whyte and colleagues that developed feeder and serum free culture conditions that took into consideration the signaling pathways necessary for avian germ cell self-renewal (Whyte et al., 2015). The work of van de Lavoie and colleagues revealed that foreign DNA can be inserted in the genome of PGCs and cells were still restricted to the germline (Van De Lavoie et al., 2006; Leighton et al., 2008). Male PGCs were cultured for a duration between 35 and 110 days during which they were transfected with a construct coding for eGFP and subsequently injected into the vasculature of White Leghorn embryos [stage 13–15 Hamburger and Hamilton (H&H)] (Van De Lavoie et al., 2006). Interestingly, the long term culture of PGCs did not influence their ability to colonize the gonads after insertion of foreign DNA, which allowed afterwards the generation of several transgenic chicken lines (Van De Lavoie et al., 2006, 2012; Macdonald et al., 2012). Leighton and colleagues gave new insights about increasing the efficiency of foreign DNA insertion in PGCs mediated by phiC31 integrase that catalyzes site-specific recombination between attB and pseudo attP sites in the chicken genome and increases transgene integration (Leighton et al., 2008).

Lu and colleagues indicated that the *piggyBac* transposon can be efficiently integrated into the genome of chicken embryo during development via electroporation (Lu et al., 2009). The transfection of PGCs with *piggyBac* transposon greatly enhanced the integration frequency of foreign DNA into the chicken genome and resulted in the generation of genetically modified chickens (Park and Han, 2012; Glover et al., 2013). In contrast, the injection of *piggyBac* transposon into the subgerminal cavity of a newly laid egg and subsequent electroporation, resulted in chickens expressing the transgene but no germline transmission was detectable (Liu et al., 2013). At the same time, Tyack and colleagues successfully developed a method for the direct transfection of circulating PGCs using Lipofectamine 2000 in combination with Tol2 transposon and transposase plasmids (Tyack et al., 2013). The plasmid contained the pCAGGS promoter driving the expression of eGFP. Tyack and colleagues found that 5/11 roosters expressed the miniTol DNA in their semen and two of them gave about 1.5% germline transmission (Tyack et al., 2013). This method substantially reduced the time needed for the *in vitro* isolation and gene manipulation of PGCs; however, it did not increase the germline transmission in G0 (Tyack et al., 2013). In addition, it does not allow clonal selection of PGCs and may result in birds with random integrations of the same transgene. Nevertheless, it is an effective method to produce genetically modified chickens as shown by various publications (Tyack et al., 2013; Lambeth et al., 2016a,b).

The possibility to culture and genetically modify chicken PGCs without losing germline competence made it possible to perform precise gene deletions and integrations in the chicken genome. Specific gene locus KO chickens were generated by Schusser and colleagues via gene targeting by homologous recombination in chicken PGCs (Schusser et al., 2013a, 2016). In the case of targeted immunoglobulin heavy chain J segment, a total of 7 from 27 PGC clones (28%) had a correctly targeted

event which reflected a high efficiency comparable to mouse ES cells (Schusser et al., 2013a). Similar efficiency was obtained after targeting the immunoglobulin light chain locus in chicken PGCs. After successful targeting of the immunoglobulin heavy or light chain in chicken PGCs, resulting clones were injected into H&H stage 13–15 embryos in order to generate germline chimeras. Germline transmission rates varied between 0.1 and 48% depending on the used PGC clone (Schusser et al., 2013a, 2016). Resulting homozygous immunoglobulin heavy chain J segment knockout birds showed a depletion of peripheral B cells and antibodies and were the first non-mammalian vertebrates harboring a knockout produced by homologs recombination. In order to perform gene knockouts by homologs recombination in PGCs, isogenic DNA is needed since mismatches in the homology regions are not tolerated (Schusser et al., 2016).

Since PGCs are precursors of sperm, researchers suggested that roosters could be used as recipient for exogenous transfer of genetically-modified PGCs which may improve the germline transmission rate (Trefil et al., 2017). Chicken embryos and adult roosters were chemically or physically sterilized to create a surrogate for external PGC donors (Trefil et al., 2006; Nakamura et al., 2008, 2010; Ghadimi et al., 2017). Nakamura and colleagues partially sterilized chicken embryos by injecting Busulfan into the yolk of fertile eggs before incubation; this led to a significant reduction of endogenous PGCs. Authors demonstrated that the sterilized embryos can be used for exogenous transfer of PGCs resulting in high efficiencies of germline transmission (Nakamura et al., 2008, 2010). Early experiments performed by Trefil and colleagues provided an alternative for chemical sterilization and concluded that repeated gamma irradiation leads to sterilization of roosters (Trefil et al., 2006). Performing injection of donor spermatogonial cells led to reestablishment of male function in 50% of the roosters only 5 weeks after injection (Trefil et al., 2006). Spermatogenesis was restored 4 weeks later in the case of PGC-transplantation compared to spermatogonial cells; however, PGCs exhibited higher efficiency in repopulating the seminiferous epithelium (Trefil et al., 2006, 2017). This was very beneficial in the case of transplantation of genetically modified PGCs into mature roosters after complete irradiation (Trefil et al., 2017). Male fertility was reestablished after the transplantation of GFP- or mCherry-expressing PGCs and resulted in almost 100% germline transmission (Trefil et al., 2017). The prominent advantage of this method is the certainty of the germline transmission and the low number of animals used in the experiment; hence reducing time and costs for testing high number of chimeric roosters. Although using gamma irradiation to sterilize roosters was as efficient as in mice, recent findings in pigs suggested that the knockout of *NANOS2*, like in mice, results in specific germline ablation with preserved testicular development (Park et al., 2017a); therefore it was suggested that *NANOS2* KO pigs may serve as a surrogate for transplantation of donor spermatogonial cells (Park et al., 2017a). Even though the importance of *NANOS2* in the transformation of ES into germ cells is well determined, little is known about its function in chickens. The most important steps made in the process of generating genetically modified chickens are summarized in **Table 1**.

TABLE 1 | Important steps for the generation of genetically modified chickens.

| Year of publication | 1987 | 1994 | 2004 | 2006 | 2013 | 2013 | 2015 | 2016 | 2017 | 2017 |
|---------------------|---|--|---|---|--|--|---|--|---------------------------------------|---|
| Contribution | Generation of the first genetically modified chicken using retroviral vectors | Generation of the first genetically modified chicken by DNA microinjection | Transgenic chickens using lentivirus carrying a transgene | Isolation and long term culture of PGCs followed by transfection and generation of eGFP transgenic chickens | Targeted gene editing in PGCs and generation of the first gene knockout chickens | <i>In vivo</i> transfection of PGCs using Lipofectamine 2000 complexed with Tol2 transposon and transposase plasmids | Development of feeder and serum-free PGC cultures | CRISPR-mediated homology directed repair targeting the immunoglobulin heavy chain locus in PGCs. | TALEN-mediated gene targeting of PGCs | Restoration of male fertility by transplantation of genetically modified PGCs |
| References | Salter et al., 1987 | Love et al., 1994 | McGrew et al., 2004 | Van De Lavoie et al., 2006 | Schusser et al., 2013a | Tyack et al., 2013 | Whyte et al., 2015 | Dimitrov et al., 2016 | Taylor et al., 2017 | Trefil et al., 2017 |

GENE EDITING IN AVIAN CELL LINES

The unavailability of fully transgenic chickens for a long time encouraged the development of alternative methods based on *in vitro* cell culture systems. *In vitro* studies helped to provide valuable data regarding host susceptibility to specific pathogens and the role of specific genes during host-pathogen interactions. DT-40 cells, an avian leukosis virus induced bursal B- cell lymphoma line, was extensively used to investigate B cell immunology, cell cycle regulation, gene conversion and apoptosis (Uckun et al., 1996; Arakawa et al., 2001; Harris et al., 2002; Arakawa and Buerstedde, 2004). A large number of DT-40 mutants were generated to understand B cell biology and were reviewed elsewhere (Arakawa and Buerstedde, 2004). For instance, studies based on DT-40 cells proved that the activation-induced cytidine deaminase (AID) triggers immunoglobulin gene diversification by gene conversion (Buerstedde et al., 1990; Kim et al., 1990). Furthermore, Szüts and colleagues used mutant DT-40 cells to demonstrate the role of RAD18 in DNA repair and the completion of gene conversion (Szüts et al., 2006). Interestingly, Schusser and colleagues replaced the immunoglobulin light and heavy chain loci in DT-40 cells with human immunoglobulin light and heavy chain loci; this led to the expression of chimeric IgM with human variable regions and chicken constant regions (Schusser et al., 2013b). The later cell line provides a model to study the diversification of the human variable region by gene conversion and somatic hypermutations in chickens. Antigen receptor analysis were performed by deep sequencing confirming that the host machinery in DT-40 cells diversified the integrated human V genes (Leighton et al., 2015).

A different established model for examining gene function in chickens is the Douglas Foster (DF-1) cells, an immortalized chicken fibroblast cell line (Foster, 1998). Recent studies used DF-1 cells to investigate host-pathogen interactions of several avian pathogens with the avian host; this included influenza A viruses, Newcastle disease virus, infectious bursal disease virus and retroviruses (Huang et al., 2003; Lee et al., 2008; Cheng et al., 2015; Hui and Leung, 2015). The overexpression of different avian genes in DF-1 cells helped to examine their role in the innate immunity against viral pathogens (Shao et al., 2014; Cheng et al., 2015; Xu et al., 2015). A well-known tool for the overexpression of various genes is the retroviral vectors derived from the SR-A strain of Rous sarcoma virus (RCAS). The RCAS system is known for its stable transduction in developing chicken embryo and cell culture (Fekete and Cepko, 1993; Bell and Brickell, 1997). Reuter and colleagues used DF-1 cells for the overexpression of the chicken IFN- α and IFN- λ (Reuter et al., 2014). The overexpression of IFN- λ in DF-1 cells did not cause substantial viral resistance against influenza A viruses H1N1, H7N1, and vesicular stomatitis virus (VSV) (Reuter et al., 2014) which suggested that DF-1 cells have weak antiviral activity of IFN- λ (Karpala et al., 2008). This was not the case for IFN- α where the overexpression led to protection against previously mentioned viruses (Reuter et al., 2014). In addition, DF-1 cells were useful to study the function of foreign genes in chicken including intracellular pattern recognition receptor such as the retinoic inducible resistant gene (RIG-I). RIG-I from duck

and goose was overexpressed in DF-1 cells and its protective effect against influenza A viruses and infectious bursitis virus (IBDV) was investigated (Barber et al., 2010; Sun et al., 2013; Shao et al., 2014). The overexpression of duck RIG-I in DF-1 cells reduced viral replication and upregulated virus-induced apoptosis following IBVD- and H9N2 influenza virus infections (Shao et al., 2014). Interestingly, the knockdown of the chicken ANP32A, a nuclear protein implicated in mRNA transport and cell death (Reilly et al., 2014), reduced the activity of different avian influenza polymerases in DF-1 cells. This indicated that avian influenza virus polymerases are more adapted to avian ANP32A and proposed this gene as target for antiviral drugs (Long et al., 2016). Furthermore, the overexpression of the chicken GADD45 β , a protein associated with cell growth control, apoptotic cell death, and the cellular response to DNA damage (Zazzeroni et al., 2003), helped to limit viral infection which could be used in the future as potential treatment for avian leukosis virus (ALV)-J infections (Zhang et al., 2016).

ALV is one of the most commonly occurring retroviruses in chickens. It induces a variety of neoplastic lesions causing losses in the productivity of affected chicken flocks (Fadly, 2000). Maas and colleagues confirmed that DF-1 cells are much more suitable than primary chicken fibroblasts (CEFs) to study host-pathogen interactions of leukosis viruses with avian cells (Maas et al., 2006). ALV was detected earlier in DF-1 cells and the infection was associated with apparent cytopathogenic effect (CPE) compared to infected-CEFs that had no apparent CPE (Maas et al., 2006). Mutations responsible for the inhibition of ALV subgroup A cell-entry were identified (Klucking et al., 2002) and consisted of four base pairs insertion and one base pair substitution in tumor virus locus A (*tva*) (Klucking et al., 2002). On the other side, only one base pair substitution in the cysteine-rich domain (CRD) of *tvb* receptor led to reduced susceptibility of DF-1 cells to infection with ALV subgroup B (Klucking et al., 2002; Reinisová et al., 2008). Interestingly, subgroup J ALV (ALV-J) uses the multimembrane-spanning cell surface protein, the chicken Na⁺/H⁺ exchanger type 1 (NHE1), as a receptor. The attachment of the virus to the receptor is crucial to initiate the infection (Barnard et al., 2006). Kučerová and colleagues used mutagenesis to introduce changes in the subgenic fragment of NHE1 (Kučerová et al., 2013); authors described the functional importance of tryptophan residue at position 38 (Trp38) for virus entry (Kučerová et al., 2013).

The rapid development of gene editing tools such as CRISPR/Cas9 rendered cell culture systems much more useful by easily targeting different genes. Precise gene editing of the chicken *NHE1* gene using CRISPR/Cas9 system led to resistance of DF-1 cells against ALV-J infection (Lee et al., 2017a). The precise genome editing of NHE1 was performed via homologs directed repair (HDR) that combined CRISPR/Cas9 vectors with single-stranded oligodeoxynucleotide (ssODNs). Authors confirmed previous observations mentioning that mutation in the Trp38 are detrimental for ALV-J infection (Kučerová et al., 2013). On the other side, non-homologous end joining repair (NHEJ) was also established in DF-1 cells. Targeting the tumor virus locus B gene, which serves as entry receptor for ALV subgroup B, resulted in frameshift mutations leading

to a KO of the *tvb*-receptor in DF-1 cells, which conferred resistance against ALV-B (Lee et al., 2017b). Abu-Bonsrah and colleagues targeted a wide range of genes in DF-1 cells such as *DROSHA*, *DICER*, *MBD3*, *KIAA1279*, *CDKN1B*, *EZH2*, *HIRA*, *TYRP1*, *STMN2*, *RET*, and *DGCR*, that play a role in embryonic development and pathogenesis of embryonic diseases (Abu-Bonsrah et al., 2016). Efficiency of inducing mutations was analyzed by T7E1 assay. The efficiency ranged between 20 and 65% in DF-1 cells (Abu-Bonsrah et al., 2016). Similar results were obtained after knocking out *KIAA1279*- and *CDKN1B*-genes in DT-40 cell-line via electroporation (Abu-Bonsrah et al., 2016). Likewise, Bai and colleagues gave more insights about the efficiency of CRISPR/Cas9 in DF-1 cells by studying gene editing in the presence and the absence of puromycin antibiotic selection (Bai et al., 2016). Three genes including peroxisome proliferator-activated receptor- γ (*PPAR- γ*), ATP synthase epsilon subunit (*ATP5E*), and ovalbumin (*OVA*) were targeted with CRISPR/Cas9 vectors. T7E assay indicated that puromycin selection increased mutation rate in the previously mentioned genes from 0.75, 0.5, and 3.0%, to 60.7, 61.3, and 47.3%, respectively (Bai et al., 2016).

GENE EDITING IN THE CHICKEN EMBRYO

The chicken embryo is a well-established model to study developmental processes, gene functions and host-pathogen interactions (Darnell and Schoenwolf, 2000; Chesnutt and Niswander, 2004; Schecterson et al., 2012). Over the last decades, different methods were established to genetically manipulate chicken embryos including electroporation of foreign DNA constructs, transduction with retroviruses and recently the combination of previous known methods with CRISPR/Cas9 system (Gandhi et al., 2017).

For example, Luo and colleagues established a protocol based on *ex ovo* electroporation of 3.5 days old chicken embryos for the overexpression of Cad7 and eGFP (Luo et al., 2012). This method provided accessibility of different embryonic parts for the electroporation, which are not easily reachable when the embryo is still inside the egg (Luo et al., 2012). Similarly, *in ovo* electroporation of the embryonic auditory brainstem was previously established (Lu et al., 2017). Plasmids of interest were successfully integrated into the nucleus magnocellularis and nucleus laminaris. Authors indicated the possibility of drug inducible gene expression which was confirmed in the presence of doxycycline (Lu et al., 2017).

A well-established tool for foreign DNA integration is the RCAS-system. Using RCAS in the chicken embryo model indicated that vector proteins and inserted transgenes were mainly detectable in the skin, blood vessels and heart (Sato et al., 2002; Kothlow et al., 2010). Several studies deduced the efficacy of RCAS-system in the case of foreign DNA integration and gene overexpression in chicken embryos (Bell and Brickell, 1997; Sato et al., 2002; Kothlow et al., 2010; Schusser et al., 2011; Reuter et al., 2014). This system is very useful to study the specific function of relevant genes for the innate immunity, particularly during the interaction with influenza A viruses. RCAS vectors expressing

various *Mx* gene isoforms were used for transduction of CEFs. Four days post-transfection, CEFs expressing the retrovirally transduced *Mx* proteins were injected in the yolk sac of 3 days-old fertilized eggs (Schusser et al., 2011). The overexpression of different *Mx* isoforms in embryonated eggs did not protect against influenza A virus infection, which was in agreement with the results obtained from chicken fibroblasts (Schusser et al., 2011). In addition, the role of IFN- λ was previously investigated by the generation of mosaic chicken embryos overexpressing chicken IFN- λ (Reuter et al., 2014). Generated embryos exhibited lower viral titers upon challenge with influenza A viruses, NDV Herts-33, or IBV M-41 via the allantoic cavity by at least four log₁₀ units compared to inoculated eggs with empty RCAS vector (Reuter et al., 2014). This clearly demonstrated the protective effect of chicken IFN- λ against different viruses (Reuter et al., 2014). Although the IFN- λ overexpression had detrimental effects at early hours post hatch (Reuter et al., 2014), RCAS system was shown to be successful for maintaining transgene expression after hatch (Kothlow et al., 2010).

A similar system based on gene transfer mediated by lentiviral vectors was described in embryonated eggs (Hen et al., 2012). Usefulness of lentiviral vectors in developmental biology was previously reviewed elsewhere (Stern, 2004). Lentiviral vectors of feline immunodeficiency virus origin were injected into chorioallantoic membrane (CAM) of 11 days old chicken embryos. The injected lentiviral vectors carried yellow fluorescent protein (YFP) or recombinant alpha-melanocyte-stimulating hormone (α -MSH) genes and they were expressed under the cytomegalovirus (CMV) promoter (Hen et al., 2012). High efficiency of transduction was observed in the liver, which implied that this model could be useful for the study of hormones and enzymes.

The application of gene editing technologies via *in ovo* electroporation of chicken embryos seems to be efficient (Wilson and Stoeckli, 2012). Wilson and Stoeckli used miRNA-based plasmids for knocking down gene expression in the chicken neural tube (Wilson and Stoeckli, 2012). Additionally, Ghandi and colleagues used *ex ovo* electroporation to knockout Pax7 and Sox10, a key transcription factors in the neural crest, leading to loss of their proteins and transcripts (Gandhi et al., 2017). Overall, collected data indicated that *in ovo* gene manipulation of the chicken embryo could be used as a model for the study of different embryonic developmental stages (Gandhi et al., 2017; Lu et al., 2017). High targeting efficiency and the simplicity of CRISPR/Cas9 make it now possible to knockout genes in specific tissues/organs of the developing chicken embryo. This allows the study the gene function during development without generating fully gene edited chicken lines.

GENERATION OF GENETICALLY MODIFIED CHICKENS

The generation of genetically modified chickens has wide applications in agricultural and biomedical research (Sang, 1994; Ivarie, 2003; Mozdziaik and Petite, 2004). Benefiting from gene editing technologies and germline transmission of PGCs, new

knowledge was brought to light about specific gene functions (Schusser et al., 2013a, 2016), resistant for infectious diseases (Lyll et al., 2011) and the possible preservation of endangered species including the Houbara Bastard (Kang et al., 2008; Wernery et al., 2010; Van De Lavoie et al., 2012). Different methods used for gene editing in chickens and the generated chicken lines were stated earlier in this review. In addition, the worldwide availability of genetically modified chicken lines is summarized in **Table 2**.

Specific gene editing in PGCs was improved using TALEN and CRISPR/Cas9 via HDR (Dimitrov et al., 2016; Oishi et al., 2016; Taylor et al., 2017). Using CRISPR/Cas9, the efficiency of gene targeting was increased remarkably in PGCs (Dimitrov et al., 2016). In order to introduce a loxP site into the immunoglobulin heavy chain locus, Dimitrov and colleagues combined a targeting vector having a total of 2 kb homology arms with CRISPR/Cas9 system targeting the upstream region of the single immunoglobulin heavy chain variable region (VH) in PGCs (Dimitrov et al., 2016). Interestingly, all selected drug resistant PGC clones contained the correct targeting event and the germline transmission rate varied between 0 and 100% depending on the used PGC line (Dimitrov et al., 2016).

Targeting the *DDX4* locus, located on the Z chromosome, showed possible role of this gene in the formation of the germ cell lineage (Taylor et al., 2017). Targeted *DDX4* KO was achieved with TALEN in combination with a targeting vector. Authors reported a germline transmission rate of 6% from the founder birds (Taylor et al., 2017). G1 female chicks were hemizygous mutant for *DDX4*, they did not lay eggs and had no yellow or white follicles in the ovaries. Surprisingly this was not the case in *DDX4* knockout female mice (Tanaka et al., 2000).

Overall, a significant progress was made in the last decade in producing and using genetically modified chickens to understand developmental biology, immunology, host-pathogen interaction, reproductive biology and physiology. However, efforts to generate resistant chickens for specific pathogens are still at the beginning, probably due to the lack of specific gene targets responsible for acquiring resistance against specific pathogens. This was not the case in other livestock including pigs which were genetically edited to gain resistance against porcine reproductive and respiratory syndrome virus (PRRSV) (Whitworth et al., 2015; Burkard et al., 2017). Using NHEJ, Whitworth and colleagues generated KO pigs with premature stop codon in exon 3 of the viral receptor CD163 (Whitworth et al., 2015). CD163-KO pigs challenged with PRRSV did not exhibit any clinical symptoms, lung pathology, viremia, or antibody response. In addition, Burkard and colleagues generated an exon 7 deletion in CD163 using two sgRNAs to induce the excision of the exon (Burkard et al., 2017). Pigs carrying the mutation were healthy and kept the main biological functions of the protein while macrophages isolated from the CD163 KO animals indicated an inhibition of the viral infection (Burkard et al., 2017).

So far, only few reports are available about the resistance of gene-edited chickens for specific pathogens. Lyall and colleagues generated transgenic chickens expressing short-hairpin RNA intended to function as a decoy that interacts and blocks influenza A virus polymerase (Lyall et al., 2011). Although birds

TABLE 2 | Worldwide availability of genetically modified chickens.

| Transgenic chicken | Affiliation of the research group | Country | References |
|---|--|----------------|---|
| Transgenic chickens carrying a benign defective subgroup A leukosis virus | Avian Disease and Oncology Laboratory, USDA Agriculture Research Service | USA | Salter and Crittenden, 1989; Cao et al., 2015 |
| Transgenic chickens expressing active β -lactamase in the egg white | AviGenics, Inc., Georgia BioBusiness Center, Athens | USA | Harvey et al., 2002 |
| eGFP expressing chickens | The Roslin Institute and Royal Dick School of Veterinary Studies, University of Edinburgh | UK | McGrew et al., 2004 |
| | Department of Physiology, Catholic University of Daegu School of Medicine | South Korea | Kwon et al., 2004 |
| | University of Utah School of Medicine, Department of Neurobiology, and Anatomy | USA | Chapman et al., 2005 |
| | Crystal Bioscience/ Ligand Pharmaceuticals Inc | USA | Van De Lavoie et al., 2006 |
| | Technical University Munich, School of Life Sciences Weihenstephan; Department of Animal Sciences, Reproductive Biotechnology | Germany | Trefil et al., 2017 |
| Hens specifically expressing therapeutic proteins in the oviduct | The Roslin Institute and Royal Dick School of Veterinary Studies, University of Edinburgh | UK | Lillico et al., 2007 |
| Production of transgenic chickens expressing a tetracycline-inducible eGFP gene | Department of Physiology, Catholic University of Daegu School of Medicine, Daegu | South Korea | Kwon et al., 2011 |
| Short-hairpin RNA against Influenza expressing chickens | Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge | UK | Lyall et al., 2011 |
| Transgenic chickens expressing human extracellular superoxide dismutase | Laboratory of Dermatology-immunology, The Catholic University of Korea | South Korea | Byun et al., 2013 |
| Immunoglobulin heavy chain (JH) KO chickens | Crystal Bioscience/ Ligand Pharmaceuticals Inc | USA | Schusser et al., 2013a |
| Transgenic chickens expressing the human urokinase type-plasminogen activator | Department of Animal Biotechnology, Bio-Organ Research Center, Konkuk University | South Korea | Lee et al., 2013 |
| CSF1R-receptor reporter chickens | The Roslin Institute and Royal Dick School of Veterinary Studies, University of Edinburgh | UK | Balic et al., 2014 |
| Immunoglobulin light chain (IgL) KO chickens | Crystal Bioscience/Ligand Pharmaceuticals Inc | USA | Schusser et al., 2016 |
| Cre-recombinase expressing chickens | Crystal Bioscience/Ligand Pharmaceuticals Inc | USA | Leighton et al., 2016 |
| Ovalbumin and Ovomucoid KO chickens | Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology and Animal Breeding and Reproduction Research Division | Japan | Oishi et al., 2016 |
| Aromatase overexpressing chickens | Department of Anatomy and Developmental Biology, Monash University, Clayton | Australia | Lambeth et al., 2016b |
| mCherry expressing chickens | Technical University Munich, School of Life Sciences Weihenstephan, Department of Animal Sciences, Reproductive Biotechnology | Germany | Trefil et al., 2017 |
| | BIOPHARM, Research Institute of Biopharmacy and Veterinary Drugs and Institute of Molecular Genetics of the Czech Academy of Sciences | Czech Republic | |
| DDX4 KO chickens | The Roslin Institute and Royal Dick School of Veterinary Studies, University of Edinburgh | UK | Taylor et al., 2017 |
| 3D8 single chain variable fragment (scFv) expressing chickens | Animal Biotechnology Division, National Institute of Animal Science and Department of Avian Disease Laboratory, College of Veterinary Medicine | South Korea | June Byun et al., 2017 |
| Chickens with humanized immunoglobulin genes | Crystal Bioscience/ Ligand Pharmaceuticals Inc | USA | Ching et al., 2018 |
| Chickens overexpressing human IFN- β | Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology | Japan | Oishi et al., 2018 |

were not resistant to initial infection, viral transmission was prevented (Lyall et al., 2011). A different study demonstrated the possibility to suppress influenza A virus transmission

in transgenic birds expressing the 3D8 single chain variable fragment (scFv), a gene that interacts with viral genome leading to suppression of viral shedding (June Byun et al., 2017).

FURTHER APPLICATIONS IN BIOMEDICAL RESEARCH

The chicken became a very interesting model in biomedical research. Different temporal patterns of bright light were used to study the effect on myopia in chickens. Lan and colleagues found that intermittent episodes of light suppress myopia in chickens more than continuous bright light (Lan et al., 2014). Although the obtained results may not be directly translated into humans (Lan et al., 2014), future applications in optical research seem to be promising. In addition, the chicken was used as a model for xenotransplantation by injection of human stem cells into small induced lesions in the chicken embryo neural tube (Boulland et al., 2010). Authors stated that the reduced immune response during early embryonic development helps to study xenotransplantation without the risk of early immune rejection (Boulland et al., 2010). The chicken was also used as a human multiple myeloma xenograft model (Martowicz et al., 2015); it was suggested that this model may offer novel therapeutic compounds targeting survival and proliferation of multiple myeloma cells. Using the chicken as a bioreactor may greatly benefit human health by providing alternative therapeutic approaches (Zhu et al., 2005). A promising approach using chickens for the production of human antibodies is the replacement of the chicken immunoglobulin variable regions by human V regions and synthetic pseudogene arrays in order to produce affinity matured human antibodies in chickens (Ching et al., 2018). The OmniChicken by Ligand Pharmaceuticals Inc. is a worldwide unique platform to produce human monoclonal antibodies from chickens making use of the phylogenetic difference between mammals and birds. The purification of overexpressed human antibodies from the chicken egg seems also to be a valid application which was reviewed elsewhere (Flemming, 2005). A very recent study conducted by Oishi and colleagues demonstrated the ability of integrating human interferon beta (hIFN- β) into the chicken ovalbumin locus in order to produce hIFN- β in egg white (Oishi et al., 2018). Authors demonstrated the ability of producing foreign proteins in eggs which would have industrial and therapeutic applications.

FUTURE PERSPECTIVES

The role of host genes in the susceptibility of chickens to different pathogens was mostly investigated *in vitro*. Preliminary *in vitro* investigations provide solid information about the role of these genes prior to the generation of fully gene edited chickens. New

technologies including CRISPR/Cas9 make the process of gene editing easy and highly efficient in contrast to the well-established process of homologs recombination. Although gene editing in mammals, particularly mice and pigs, is vastly advanced, gene editing in chickens is entering the golden age. For instance the generation of Cas9-expressing pigs will provide a powerful tool for the study of biological processes (Wang et al., 2017); while this was not done yet in chickens, it seems to be beneficial and may be used in the future to dissect unknown gene functions faster and more easily.

Therapeutic applications using human monoclonal antibodies produced from humanized chickens may be beneficial over *in vitro* approaches lacking affinity maturation (Ching et al., 2018). In addition, production of antibodies in chicken eggs represents an economic and stress-free method for the production of specific antibodies (Amro et al., 2018). Using chicken eggs to manufacture specific proteins in eggs seems interesting (Lillico et al., 2005; Petitte and Mozdziak, 2007) especially since it may allow improvement of digestibility of sugar complexes in feedstuffs; however, this application may be thwarted by critics that claim the inedibility of the product.

Several advantages are provided by newly invented gene editing technologies including the simplicity of design and application combined with high efficiency (Chira et al., 2017). Understanding the host cell behavior during host-pathogen interactions may help targeting pathogen specific receptors and viral cellular transport (Heaton et al., 2016). The determination of new target genes associated with disease susceptibility should fill the research gap and open the door for new therapeutical approaches. Although the debate about using genetically modified animals in food production will continue to be stimulated, we may obtain new breeds of chickens in the future that are resistant for specific pathogens. We speculate that spending more efforts connecting gene editing technologies with the prevention of infectious diseases will change the way we use to fight pathogens and will probably improve the animal welfare.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was supported by grant Schu2446/3-1 to BS from the Deutsche Forschungsgemeinschaft.

REFERENCES

- Abu-Bonsrah, K. D., Zhang, D., and Newgreen, D. F. (2016). CRISPR/Cas9 targets chicken embryonic somatic cells *in vitro* and *in vivo* and generates phenotypic abnormalities. *Sci. Rep.* 6:34524. doi: 10.1038/srep34524
- Amro, W. A., Al-Qaisi, W., and Al-Razem, F. (2018). Production and purification of IgY antibodies from chicken egg yolk. *J. Genet. Eng. Biotechnol.* 16, 99–103. doi: 10.1016/j.jgeb.2017.10.003
- Arakawa, H., and Buerstedde, J. M. (2004). Immunoglobulin gene conversion: insights from bursal B cells and the DT40 cell line. *Dev. Dyn.* 229, 458–464. doi: 10.1002/dvdy.10495
- Arakawa, H., Lodygin, D., and Buerstedde, J. M. (2001). Mutant loxP vectors for selectable marker recycle and conditional knock-outs. *BMC Biotechnol.* 1:7. doi: 10.1186/1472-6750-1-7
- Bai, Y., He, L., Li, P., Xu, K., Shao, S., Ren, C., et al. (2016). Efficient genome editing in chicken DF-1 cells using the CRISPR/Cas9 system. *G3: Genes Genomes Genet.* 6, 917–923. doi: 10.1534/g3.116.027706

- Balic, A., Garcia-Morales, C., Vervelde, L., Gilhooley, H., Sherman, A., Garceau, V., et al. (2014). Visualisation of chicken macrophages using transgenic reporter genes: insights into the development of the avian macrophage lineage. *Development* 141, 3255–3265. doi: 10.1242/dev.105593
- Barber, M. R., Aldridge, J. R., Webster, R. G., and Magor, K. E. (2010). Association of RIG-I with innate immunity of ducks to influenza. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5913–5918. doi: 10.1073/pnas.1001755107
- Barnard, R. J., Elleder, D., and Young, J. A. (2006). Avian sarcoma and leukosis virus-receptor interactions: from classical genetics to novel insights into virus-cell membrane fusion. *Virology* 344, 25–29. doi: 10.1016/j.virol.2005.09.021
- Bell, E. J., and Brickell, P. M. (1997). Replication-competent retroviral vectors for expressing genes in avian cells *in vitro* and *in vivo*. *Mol. Biotechnol.* 7, 289–298. doi: 10.1007/BF02740819
- Bosselman, R. A., Hsu, R. Y., Boggs, T., Hu, S., Bruszewski, J., Ou, S., et al. (1989). Germline transmission of exogenous genes in the chicken. *Science* 243, 533–535. doi: 10.1126/science.2536194
- Boulland, J.-L., Halasi, G., Kasumacic, N., and Glover, J. C. (2010). Xenotransplantation of human stem cells into the chicken embryo. *J. Vis. Exp.* 11:2071. doi: 10.3791/2071
- Bryan, F. L., and Doyle, M. P. (1995). Health risks and consequences of *Salmonella* and *Campylobacter jejuni* in raw poultry. *J. Food Prot.* 58, 326–344. doi: 10.4315/0362-028X-58.3.326
- Buerstedde, J. M., Reynaud, C. A., Humphries, E. H., Olson, W., Ewert, D. L., and Weill, J. C. (1990). Light chain gene conversion continues at high rate in an ALV-induced cell line. *EMBO J.* 9, 921–927. doi: 10.1002/j.1460-2075.1990.tb08190.x
- Burkard, C., Lillico, S. G., Reid, E., Jackson, B., Mileham, A. J., Ait-Ali, T., et al. (2017). Precision engineering for PRRSV resistance in pigs: macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function. *PLoS Pathog.* 13:e1006206. doi: 10.1371/journal.ppat.1006206
- Byun, S. J., Ji, M.-R., Jang, Y.-J., Hwang, A.-I., Chung, H. K., Kim, J. S., et al. (2013). Human extracellular superoxide dismutase (EC-SOD) expression in transgenic chicken. *BMB Rep.* 46, 404–409. doi: 10.5483/BMBRep.2013.46.8.251
- Cao, W., Mays, J., Kulkarni, G., Dunn, J., Fulton, R. M., and Fadly, A. (2015). Further observations on serotype 2 Marek's disease virus-induced enhancement of spontaneous avian leukosis virus-like bursal lymphomas in ALVA6 transgenic chickens. *Avian Pathol.* 44, 23–27. doi: 10.1080/03079457.2014.989195
- Carlson, D. F., Tan, W., Lillico, S. G., Stverakova, D., Proudfoot, C., Christian, M., et al. (2012). Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17382–17387. doi: 10.1073/pnas.1211446109
- Chapman, S. C., Lawson, A., Macarthur, W. C., Wiese, R. J., Loechel, R. H., Burgos-Trinidad, M., et al. (2005). Ubiquitous GFP expression in transgenic chickens using a lentiviral vector. *Development* 132, 935–940. doi: 10.1242/dev.01652
- Cheng, Y., Sun, Y., Wang, H., Yan, Y., Ding, C., and Sun, J. (2015). Chicken STING mediates activation of the IFN gene independently of the RIG-I gene. *J. Immunol.* 195, 3922–3936. doi: 10.4049/jimmunol.1500638
- Chesnutt, C., and Niswander, L. (2004). Plasmid-based short-hairpin RNA interference in the chicken embryo. *Genesis* 39, 73–78. doi: 10.1002/gene.20028
- Ching, K. H., Collarini, E. J., Abdiche, Y. N., Bedinger, D., Pedersen, D., Izquierdo, S., et al. (2018). Chickens with humanized immunoglobulin genes generate antibodies with high affinity and broad epitope coverage to conserved targets. *MAbs* 10, 71–80. doi: 10.1080/19420862.2017.1386825
- Chira, S., Gulei, D., Hajitou, A., Zimta, A.-A., Cordelier, P., and Berindan-Neagoe, I. (2017). CRISPR/Cas9: transcending the reality of genome editing. *Molecul. Ther. Nucleic Acids* 7, 211–222. doi: 10.1016/j.omtn.2017.04.001
- Costantini, F., and Lacy, E. (1981). Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature* 294, 92–94. doi: 10.1038/294092a0
- Darnell, D. K., and Schoenwolf, G. C. (2000). The chick embryo as a model system for analyzing mechanisms of development. *Dev. Biol. Protocol.* 135, 25–29. doi: 10.1385/1-59259-685-1:25
- Dimitrov, L., Pedersen, D., Ching, K. H., Yi, H., Collarini, E. J., Izquierdo, S., et al. (2016). Germline gene editing in chickens by efficient CRISPR-mediated homologous recombination in primordial germ cells. *PLoS ONE* 11:e0154303. doi: 10.1371/journal.pone.0154303
- Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292, 154–156. doi: 10.1038/292154a0
- Fadly, A. M. (2000). Isolation and identification of avian leukosis viruses: a review. *Avian Pathol.* 29, 529–535. doi: 10.1080/03079450020016760
- Fekete, D. M., and Cepko, C. L. (1993). Replication-competent retroviral vectors encoding alkaline phosphatase reveal spatial restriction of viral gene expression/transduction in the chick embryo. *Mol. Cell. Biol.* 13, 2604–2613. doi: 10.1128/MCB.13.4.2604
- Flemming, A. (2005). Biotechnology: human antibodies from chicken eggs. *Nat. Rev. Drug Discov.* 4:884. doi: 10.1038/nrd1883
- Foster, D. (1998). Development of a spontaneously immortalized chicken embryo fibroblastic cell line. *Virology* 248, 305–311.
- Fujimoto, T., Ninomiya, T., and Ukeshima, A. (1976). Observations of the primordial germ cells in blood samples from the chick embryo. *Dev. Biol.* 49, 278–282. doi: 10.1016/0012-1606(76)90273-6
- Gaj, T., Gersbach, C. A., and Barbas Iii, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405. doi: 10.1016/j.tibtech.2013.04.004
- Gandhi, S., Piacentino, M. L., Viece, F. M., and Bronner, M. E. (2017). Optimization of CRISPR/Cas9 genome editing for loss-of-function in the early chick embryo. *Dev. Biol.* 432, 86–97. doi: 10.1016/j.ydbio.2017.08.036
- Gao, R., Cao, B., Hu, Y., Feng, Z., Wang, D., Hu, W., et al. (2013). Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* 368, 1888–1897. doi: 10.1056/NEJMoa1304459
- Ghadimi, F., Shakeri, M., Zhandi, M., Zaghari, M., Piryaei, A., Moslehifard, P., et al. (2017). Different approaches to establish infertile rooster. *Anim. Reprod. Sci.* 186, 31–36. doi: 10.1016/j.anireprosci.2017.08.019
- Glover, J. D., Taylor, L., Sherman, A., Zeiger-Poli, C., Sang, H. M., and McGrew, M. J. (2013). A novel piggyBac transposon inducible expression system identifies a role for AKT signalling in primordial germ cell migration. *PLoS ONE* 8:e77222. doi: 10.1371/journal.pone.0077222
- Gordon, J. W., and Ruddle, F. H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214, 1244–1246. doi: 10.1126/science.6272397
- Hammer, R. E., Pursell, V. G., Rexroad C. E. Jr, Wall, R. J., Bolt, D. J., Ebert, K. M., et al. (1985). Production of transgenic rabbits, sheep and pigs by microinjection. *Nature* 315, 680–683. doi: 10.1038/315680a0
- Harris, R. S., Sale, J. E., Petersen-Mahrt, S. K., and Neuberger, M. S. (2002). AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr. Biol.* 12, 435–438. doi: 10.1016/S0960-9822(02)00717-0
- Harvey, A. J., and Ivarie, R. (2003). Validating the hen as a bioreactor for the production of exogenous proteins in egg white. *Poult. Sci.* 82, 927–930. doi: 10.1093/ps/82.6.927
- Harvey, A. J., Speksnijder, G., Baugh, L. R., Morris, J. A., and Ivarie, R. (2002). Expression of exogenous protein in the egg white of transgenic chickens. *Nat. Biotechnol.* 20:396. doi: 10.1038/nbt0402-396
- Heaton, N. S., Moshkina, N., Fenouil, R., Gardner, T. J., Aguirre, S., Shah, P. S., et al. (2016). Targeting viral proteostasis limits influenza virus, HIV, and dengue virus infection. *Immunity* 44, 46–58. doi: 10.1016/j.immuni.2015.12.017
- Hen, G., Yosefi, S., Shinder, D., Or, A., Mygdal, S., Condiotti, R., et al. (2012). Gene transfer to chicks using lentiviral vectors administered via the embryonic chorioallantoic membrane. *PLoS ONE* 7:e36531. doi: 10.1371/journal.pone.0036531
- Houdebine, L.-M. (2002). The methods to generate transgenic animals and to control transgene expression. *J. Biotechnol.* 98, 145–160. doi: 10.1016/S0168-1656(02)00129-3
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278. doi: 10.1016/j.cell.2014.05.010
- Huang, Z., Krishnamurthy, S., Panda, A., and Samal, S. K. (2003). Newcastle disease virus V protein is associated with viral pathogenesis and functions as an alpha interferon antagonist. *J. Virol.* 77, 8676–8685. doi: 10.1128/JVI.77.16.8676-8685.2003
- Hughes, S. H., Kosik, E., Fadly, A. M., Salter, D. W., and Crittenden, L. B. (1986). Design of retroviral vectors for the insertion of foreign deoxyribonucleic acid sequences into the avian germ line. *Poult. Sci.* 65, 1459–1467. doi: 10.3382/ps.0651459

- Hui, R. K., and Leung, F. C. (2015). Differential expression profile of chicken embryo fibroblast DF-1 cells infected with cell-adapted infectious bursal disease virus. *PLoS ONE* 10:e0111771. doi: 10.1371/journal.pone.0111771
- Humphrey, T. (2006). Are happy chickens safer chickens? Poultry welfare and disease susceptibility. *Br. Poult. Sci.* 47, 379–391. doi: 10.1080/00071660600829084
- Ivarie, R. (2003). Avian transgenesis: progress towards the promise. *Trends Biotechnol.* 21, 14–19. doi: 10.1016/S0167-7799(02)00009-4
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821. doi: 10.1126/science.1225829
- June Byun, S., Yuk, S.-S., Jang, Y.-J., Choi, H., Jeon, M.-H., Erdene-Ochir, T., et al. (2017). Transgenic chickens expressing the 3D8 single chain variable fragment protein suppress avian influenza transmission. *Sci. Rep.* 7:5938. doi: 10.1038/s41598-017-05270-8
- Kanatsu-Shinohara, M., Ogonuki, N., Inoue, K., Miki, H., Ogura, A., Toyokuni, S., et al. (2003). Long-term proliferation in culture and germline transmission of mouse male germline stem cells. *Biol. Reprod.* 69, 612–616. doi: 10.1095/biolreprod.103.017012
- Kang, K. S., Lee, H. C., Kim, H. J., Lee, H. G., Kim, Y. M., Lee, H. J., et al. (2015). Spatial and temporal action of chicken primordial germ cells during initial migration. *Reproduction* 149, 179–187. doi: 10.1530/REP-14-0433
- Kang, S. J., Choi, J. W., Kim, S. Y., Park, K. J., Kim, T. M., Lee, Y. M., et al. (2008). Reproduction of wild birds via interspecies germ cell transplantation. *Biol. Reprod.* 79, 931–937. doi: 10.1095/biolreprod.108.069989
- Karpala, A. J., Morris, K. R., Broadway, M. M., Mcwaters, P. G., O'neil, T. E., Goossens, K. E., et al. (2008). Molecular cloning, expression, and characterization of chicken IFN- λ . *J. Interferon Cytokine Res.* 28, 341–350. doi: 10.1089/jir.2007.0117
- Kim, J. N., Park, T. S., Park, S. H., Park, K. J., Kim, T. M., Lee, S. K., et al. (2010). Migration and proliferation of intact and genetically modified primordial germ cells and the generation of a transgenic chicken. *Biol. Reprod.* 82, 257–262. doi: 10.1095/biolreprod.109.079723
- Kim, S., Humphries, E. H., Tjoelker, L., Carlson, L., and Thompson, C. B. (1990). Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. *Mol. Cell. Biol.* 10, 3224–3231. doi: 10.1128/MCB.10.6.3224
- Klucking, S., Adkins, H. B., and Young, J. A. (2002). Resistance to infection by subgroups B, D, and E avian sarcoma and leukosis viruses is explained by a premature stop codon within a resistance allele of the tvb receptor gene. *J. Virol.* 76, 7918–7921. doi: 10.1128/JVI.76.15.7918-7921.2002
- Kothlow, S., Schenk-Weibhauser, K., Ratcliffe, M. J., and Kaspers, B. (2010). Prolonged effect of BAFF on chicken B cell development revealed by RCAS retroviral gene transfer *in vivo*. *Mol. Immunol.* 47, 1619–1628. doi: 10.1016/j.molimm.2010.01.011
- Kučerová, D., Plachý, J., Reinišová, M., Šenigl, F., Trejbalová, K., Geryk, J., and Hejnar, J. (2013). Nonconserved tryptophan 38 of the cell surface receptor for subgroup J avian leukosis virus discriminates sensitive from resistant avian species. *J. Virol.* 87, 8399–8407. doi: 10.1128/JVI.03180-12
- Kwon, M. S., Koo, B. C., Choi, B. R., Lee, H. T., Kim, Y. H., Ryu, W.-S., et al. (2004). Development of transgenic chickens expressing enhanced green fluorescent protein. *Biochem. Biophys. Res. Commun.* 320, 442–448. doi: 10.1016/j.bbrc.2004.05.197
- Kwon, M. S., Koo, B. C., Roh, J. Y., Kim, M., Kim, J.-H., and Kim, T. (2011). Production of transgenic chickens expressing a tetracycline-inducible GFP gene. *Biochem. Biophys. Res. Commun.* 410, 890–894. doi: 10.1016/j.bbrc.2011.06.088
- Lai, L., Kolber-Simonds, D., Park, K.-W., Cheong, H.-T., Greenstein, J. L., Im, G.-S., et al. (2002). Production of α -1, 3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science* 295, 1089–1092. doi: 10.1126/science.1068228
- Lalonde, R., Fukuchi, K., and Strazielle, C. (2012). Neurologic and motor dysfunctions in APP transgenic mice. *Rev. Neurosci.* 23, 363–379. doi: 10.1515/revneuro-2012-0041
- Lambeth, L. S., Morris, K., Ayers, K. L., Wise, T. G., O'neil, T., Wilson, S., et al. (2016a). Overexpression of anti-mullerian hormone disrupts gonadal sex differentiation, blocks sex hormone synthesis, and supports cell autonomous sex development in the chicken. *Endocrinology* 157, 1258–1275. doi: 10.1210/en.2015-1571
- Lambeth, L. S., Morris, K. R., Wise, T. G., Cummins, D. M., O'neil, T. E., Cao, Y., et al. (2016b). Transgenic chickens overexpressing aromatase have high estrogen levels but maintain a predominantly male phenotype. *Endocrinology* 157, 83–90. doi: 10.1210/en.2015-1697
- Lan, W., Feldkaemper, M., and Schaeffel, F. (2014). Intermittent episodes of bright light suppress myopia in the chicken more than continuous bright light. *PLoS ONE* 9:e110906. doi: 10.1371/journal.pone.0110906
- Lee, C.-W., Jung, K., Jadhao, S. J., and Suarez, D. L. (2008). Evaluation of chicken-origin (DF-1) and quail-origin (QT-6) fibroblast cell lines for replication of avian influenza viruses. *J. Virol. Methods* 153, 22–28. doi: 10.1016/j.jviromet.2008.06.019
- Lee, H. J., Lee, K. Y., Jung, K. M., Park, K. J., Lee, K. O., Suh, J. Y., et al. (2017a). Precise gene editing of chicken Na⁺/H⁺ exchange type 1 (chNHE1) confers resistance to avian leukosis virus subgroup J (ALV-J). *Dev. Comp. Immunol.* 77, 340–349. doi: 10.1016/j.dci.2017.09.006
- Lee, H. J., Lee, K. Y., Park, Y. H., Choi, H. J., Yao, Y., Nair, V., et al. (2017b). Acquisition of resistance to avian leukosis virus subgroup B through mutations on tvb cysteine-rich domains in DF-1 chicken fibroblasts. *Vet. Res.* 48:48.
- Lee, J. H., Park, J.-W., Kim, S. W., Park, J., and Park, T. S. (2017c). CXCR4 chemokine receptor type 4 (CXCR4) is a key receptor for chicken primordial germ cell migration. *J. Reprod. Dev.* 63, 555–562. doi: 10.1262/jrd.2017-067
- Lee, S. H., Gupta, M. K., Ho, Y. T., Kim, T., and Lee, H. T. (2013). Transgenic chickens expressing human urokinase-type plasminogen activator. *Poult. Sci.* 92, 2396–2403. doi: 10.3382/ps.2013-03223
- Leighton, P. A., Pedersen, D., Ching, K., Collarini, E. J., Izquierdo, S., Jacob, R., et al. (2016). Generation of chickens expressing Cre recombinase. *Trans. Res.* 25, 609–616. doi: 10.1007/s11248-016-9952-6
- Leighton, P. A., Schusser, B., Yi, H., Glanville, J., and Harriman, W. (2015). A diverse repertoire of human immunoglobulin variable genes in a chicken B cell line is generated by both gene conversion and somatic hypermutation. *Front. Immunol.* 6:126. doi: 10.3389/fimmu.2015.00126
- Leighton, P. A., Van De Lavoie, M. C., Diamond, J. H., Xia, C., and Etches, R. J. (2008). Genetic modification of primordial germ cells by gene trapping, gene targeting, and ϕ C31 integrase. *Mol. Reprod. Dev.* 75, 1163–1175. doi: 10.1002/mrd.20859
- Lillico, S. G., McGrew, M. J., Sherman, A., and Sang, H. M. (2005). Transgenic chickens as bioreactors for protein-based drugs. *Drug Discov. Today* 10, 191–196. doi: 10.1016/S1359-6446(04)03317-3
- Lillico, S. G., Sherman, A., McGrew, M. J., Robertson, C. D., Smith, J., Haslam, C., et al. (2007). Oviduct-specific expression of two therapeutic proteins in transgenic hens. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1771–1776. doi: 10.1073/pnas.0610401104
- Liu, X., Li, N., Hu, X., Zhang, R., Li, Q., Cao, D., et al. (2013). Efficient production of transgenic chickens based on piggyBac. *Trans. Res.* 22, 417–423. doi: 10.1007/s11248-012-9642-y
- Long, J. S., Giotis, E. S., Moncorgé, O., Frise, R., Mistry, B., James, J., et al. (2016). Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature* 529, 101–104. doi: 10.1038/nature16474
- Love, J., Gribbin, C., Mather, C., and Sang, H. (1994). Transgenic birds by DNA microinjection. *Biotechnology* 12, 60–63.
- Lu, T., Cohen, A. L., and Sanchez, J. T. (2017). In Ovo Electroporation in the Chicken Auditory Brainstem. *J. Vis. Exp.* 9.
- Lu, Y., Lin, C., and Wang, X. (2009). PiggyBac transgenic strategies in the developing chicken spinal cord. *Nucleic Acids Res.* 37:e141. doi: 10.1093/nar/gkp686
- Luo, J., Yan, X., Lin, J., and Rolf, A. (2012). Gene transfer into older chicken embryos by *ex ovo* electroporation. *J. Vis. Exp.* 27:4078. doi: 10.3791/4078
- Lyall, J., Irvine, R. M., Sherman, A., McKinley, T. J., Nú-éz, A., Purdie, A., et al. (2011). Suppression of avian influenza transmission in genetically modified chickens. *Science* 331, 223–226. doi: 10.1126/science.1198020
- Maas, R., Van Zoelen, D., Oei, H., and Claassen, I. (2006). Replacement of primary chicken embryonic fibroblasts (CEF) by the DF-1 cell line for detection of avian leukosis viruses. *Biologicals* 34, 177–181. doi: 10.1016/j.biologicals.2005.09.002
- Macdonald, J., Taylor, L., Sherman, A., Kawakami, K., Takahashi, Y., Sang, H. M., et al. (2012). Efficient genetic modification and germ-line transmission of

- primordial germ cells using piggyBac and Tol2 transposons. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1466–E1472. doi: 10.1073/pnas.1118715109
- Martowicz, A., Kern, J., Gunsilius, E., and Untergasser, G. (2015). Establishment of a human multiple myeloma xenograft model in the chicken to study tumor growth, invasion and angiogenesis. *J. Vis. Exp.* 1:e52665. doi: 10.3791/52665
- McCreath, K. J., Howcroft, J., Campbell, K. H., Colman, A., Schnieke, A. E., and Kind, A. J. (2000). Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* 405, 1066–1069. doi: 10.1038/35016604
- McGrew, M. J., Sherman, A., Ellard, F. M., Lillico, S. G., Gilhooley, H. J., Kingsman, A. J., et al. (2004). Efficient production of germline transgenic chickens using lentiviral vectors. *EMBO Rep.* 5, 728–733. doi: 10.1038/sj.embor.7400171
- Mozdziak, P. E., Borwornpinyo, S., McCoy, D. W., and Petitte, J. N. (2003). Development of transgenic chickens expressing bacterial beta-galactosidase. *Dev. Dyn.* 226, 439–445. doi: 10.1002/dvdy.10234
- Mozdziak, P. E., and Petitte, J. N. (2004). Status of transgenic chicken models for developmental biology. *Dev. Dyn.* 229, 414–421. doi: 10.1002/dvdy.10461
- Nakamura, Y., Usui, F., Ono, T., Takeda, K., Nirasawa, K., Kagami, H., et al. (2010). Germline replacement by transfer of primordial germ cells into partially sterilized embryos in the chicken. *Biol. Reprod.* 83, 130–137. doi: 10.1095/biolreprod.110.083923
- Nakamura, Y., Yamamoto, Y., Usui, F., Atsumi, Y., Ito, Y., Ono, T., et al. (2008). Increased proportion of donor primordial germ cells in chimeric gonads by sterilisation of recipient embryos using busulfan sustained-release emulsion in chickens. *Reprod. Fertility Dev.* 20, 900–907. doi: 10.1071/RD08138
- Nottle, M. B., Beebe, L. F., Harrison, S. J., McIlpatrick, S. M., Ashman, R. J., O'Connell, P. J., et al. (2007). Production of homozygous α -1, 3-galactosyltransferase knockout pigs by breeding and somatic cell nuclear transfer. *Xenotransplantation* 14, 339–344. doi: 10.1111/j.1399-3089.2007.00417.x
- Oishi, I., Yoshii, K., Miyahara, D., Kagami, H., and Tagami, T. (2016). Targeted mutagenesis in chicken using CRISPR/Cas9 system. *Sci. Rep.* 6:23980. doi: 10.1038/srep23980
- Oishi, I., Yoshii, K., Miyahara, D., and Tagami, T. (2018). Efficient production of human interferon beta in the white of eggs from ovalbumin gene-targeted hens. *Sci. Rep.* 8:10203. doi: 10.1038/s41598-018-28438-2
- Park, K. E., Kaucher, A. V., Powell, A., Waqas, M. S., Sandmaier, S. E., Oatley, M. J., et al. (2017a). Generation of germline ablated male pigs by CRISPR/Cas9 editing of the NANOS2 gene. *Sci. Rep.* 7:40176. doi: 10.1038/srep40176
- Park, M. H., Lee, H. J., Lee, H. L., Son, D. J., Ju, J. H., Hyun, B. K., et al. (2017b). Parkin knockout inhibits neuronal development via regulation of proteasomal degradation of p21. *Theranostics* 7, 2033–2045. doi: 10.7150/thno.19824
- Park, T. S., and Han, J. Y. (2012). piggyBac transposition into primordial germ cells is an efficient tool for transgenesis in chickens. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9337–9341. doi: 10.1073/pnas.1203823109
- Perdue, M. L., and Swayne, D. E. (2005). Public health risk from avian influenza viruses. *Avian Dis.* 49, 317–327. doi: 10.1637/7390-060305R.1
- Perleberg, C., Kind, A., and Schnieke, A. (2018). Genetically engineered pigs as models for human disease. *Dis. Model. Mech.* 11:dmm030783. doi: 10.1242/dmm.030783
- Petitte, J. N., Clark, M. E., Liu, G., Verrinder Gibbins, A. M., and Etches, R. J. (1990). Production of somatic and germline chimeras in the chicken by transfer of early blastodermal cells. *Development* 108, 185–189.
- Petitte, J. N., and Mozdziak, P. E. (2007). The incredible, edible, and therapeutic egg. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1739–1740. doi: 10.1073/pnas.0611652104
- Pinkert, C. A. (ed.). (2014). “1 - Introduction to Transgenic Animal Technology,” in *Transgenic Animal Technology*, 3rd Edn, (London: Elsevier), 3–13.
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143
- Reilly, P. T., Yu, Y., Hamiche, A., and Wang, L. (2014). Cracking the ANP32 whips: important functions, unequal requirement, and hints at disease implications. *Bioessays* 36, 1062–1071. doi: 10.1002/bies.201400058
- Reinisová, M., Senigl, F., Yin, X., Plachy, J., Geryk, J., Elleder, D., et al. (2008). A single-amino-acid substitution in the Tvbs1 receptor results in decreased susceptibility to infection by avian sarcoma and leukosis virus subgroups B and D and resistance to infection by subgroup E *in vitro* and *in vivo*. *J. Virol.* 82, 2097–2105. doi: 10.1128/JVI.02206-07
- Reuter, A., Soubies, S., Härtle, S., Schusser, B., Kaspers, B., Staeheli, P., et al. (2014). Antiviral activity of lambda interferon in chickens. *J. Virol.* 88, 2835–2843. doi: 10.1128/JVI.02764-13
- Salter, D. W., and Crittenden, L. B. (1989). Artificial insertion of a dominant gene for resistance to avian leukosis virus into the germ line of the chicken. *Theor. Appl. Genet.* 77, 457–461. doi: 10.1007/BF00274263
- Salter, D. W., Smith, E. J., Hughes, S. H., Wright, S. E., and Crittenden, L. B. (1987). Transgenic chickens: insertion of retroviral genes into the chicken germ line. *Virology* 157, 236–240. doi: 10.1016/0042-6822(87)90334-5
- Sang, H. (1994). Transgenic chickens—methods and potential applications. *Trends Biotechnol.* 12, 415–420. doi: 10.1016/0167-7799(94)90030-2
- Sato, N., Matsuda, K., Sakuma, C., Foster, D. N., Oppenheim, R. W., and Yaginuma, H. (2002). Regulated gene expression in the chicken embryo by using replication-competent retroviral vectors. *J. Virol.* 76, 1980–1985. doi: 10.1128/JVI.76.4.1980-1985.2002
- Schecterson, L. C., Sanchez, J. T., Rubel, E. W., and Bothwell, M. (2012). TrkB downregulation is required for dendrite retraction in developing neurons of chicken nucleus magnocellularis. *J. Neurosci.* 32, 14000–14009. doi: 10.1523/JNEUROSCI.2274-12.2012
- Schnieke, A. E., Kind, A. J., Ritchie, W. A., Mycock, K., Scott, A. R., Ritchie, M., et al. (1997). Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science* 278, 2130–2133. doi: 10.1126/science.278.5346.2130
- Schusser, B., Collarini, E. J., Pedersen, D., Yi, H., Ching, K., Izquierdo, S., et al. (2016). Expression of heavy chain-only antibodies can support B-cell development in light chain knockout chickens. *Eur. J. Immunol.* 46, 2137–2148. doi: 10.1002/eji.201546171
- Schusser, B., Collarini, E. J., Yi, H., Izquierdo, S. M., Fesler, J., Pedersen, D., et al. (2013a). Immunoglobulin knockout chickens via efficient homologous recombination in primordial germ cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20170–20175. doi: 10.1073/pnas.1317106110
- Schusser, B., Reuter, A., Von Der Malsburg, A., Penski, N., Weigend, S., Kaspers, B., et al. (2011). Mx is dispensable for interferon-mediated resistance of chicken cells against influenza A virus. *J. Virol.* 85, 8307–8315. doi: 10.1128/JVI.00535-11
- Schusser, B., Yi, H., Collarini, E. J., Izquierdo, S. M., Harriman, W. D., Etches, R. J., et al. (2013b). Harnessing gene conversion in chicken B cells to create a human antibody sequence repertoire. *PLoS ONE* 8:e80108. doi: 10.1371/journal.pone.0080108
- Shao, Q., Xu, W., Yan, L., Liu, J., Rui, L., Xiao, X., et al. (2014). Function of duck RIG-I in induction of antiviral response against IBDV and avian influenza virus on chicken cells. *Virus Res.* 191, 184–191. doi: 10.1016/j.virusres.2014.07.028
- Smith, C. A., Roeszler, K. N., and Sinclair, A. H. (2009). Robust and ubiquitous GFP expression in a single generation of chicken embryos using the avian retroviral vector, RCASBP. *Differentiation* 77, 473–482. doi: 10.1016/j.diff.2009.02.001
- Soto, D. A., and Ross, P. J. (2016). Pluripotent stem cells and livestock genetic engineering. *Transgenic Res.* 25, 289–306. doi: 10.1007/s11248-016-9929-5
- Speedy, A. W. (2003). Global production and consumption of animal source foods. *J. Nutrition* 133, 4048S–4053S. doi: 10.1093/jn/133.11.4048S
- Stebler, J., Spieler, D., Slanchev, K., Molyneux, K. A., Richter, U., Cojocaru, V., et al. (2004). Primordial germ cell migration in the chick and mouse embryo: the role of the chemokine SDF-1/CXCL12. *Dev. Biol.* 272, 351–361. doi: 10.1016/j.ydbio.2004.05.009
- Stern, C. D. (1990). The marginal zone and its contribution to the hypoblast and primitive streak of the chick embryo. *Development* 109, 667–682.
- Stern, C. D. (2004). *The Chick Embryo—Past, Present and Future as a Model System in Developmental Biology*. London: Elsevier.
- Stern, C. D. (2005). The chick; a great model system becomes even greater. *Dev. Cell* 8, 9–17. doi: 10.1016/j.devcel.2004.11.018
- Sun, Y., Ding, N., Ding, S. S., Yu, S., Meng, C., Chen, H., et al. (2013). Goose RIG-I functions in innate immunity against Newcastle disease virus infections. *Mol. Immunol.* 53, 321–327. doi: 10.1016/j.molimm.2012.08.022
- Szűts, D., Simpson, L. J., Kabani, S., Yamazoe, M., and Sale, J. E. (2006). Role for RAD18 in homologous recombination in DT40 cells. *Mol. Cell. Biol.* 26, 8032–8041. doi: 10.1128/MCB.01291-06
- Talbot, N. C., and Blomberg, L. A. (2008). The pursuit of ES cell lines of domesticated ungulates. *Stem Cell Rev.* 4, 235–254. doi: 10.1007/s12015-008-9026-0

- Tanaka, S. S., Toyooka, Y., Akasu, R., Katoh-Fukui, Y., Nakahara, Y., Suzuki, R., et al. (2000). The mouse homolog of *Drosophila* Vasa is required for the development of male germ cells. *Genes Dev.* 14, 841–853. doi: 10.1101/gad.14.7.841
- Taylor, L., Carlson, D. F., Nandi, S., Sherman, A., Fahrenkrug, S. C., and McGrew, M. J. (2017). Efficient TALEN-mediated gene targeting of chicken primordial germ cells. *Development* 144, 928–934. doi: 10.1242/dev.145367
- Thomas, K. R., and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51, 503–512. doi: 10.1016/0092-8674(87)90646-5
- Trefil, P., Aumann, D., Koslová, A., Mucksová, J., Benešová, B., Kalina, J., et al. (2017). Male fertility restored by transplanting primordial germ cells into testes: a new way towards efficient transgenesis in chicken. *Sci. Rep.* 7:14246. doi: 10.1038/s41598-017-14475-w
- Trefil, P., Micáková, A., Mucksová, J., Hejnar, J., Poplstein, M., Bakst, M. R., et al. (2006). Restoration of spermatogenesis and male fertility by transplantation of dispersed testicular cells in the chicken. *Biol. Reprod.* 75, 575–581. doi: 10.1095/biolreprod.105.050278
- Tyack, S. G., Jenkins, K. A., O'neil, T. E., Wise, T. G., Morris, K. R., Bruce, M. P., et al. (2013). A new method for producing transgenic birds via direct *in vivo* transfection of primordial germ cells. *Trans. Res.* 22, 1257–1264. doi: 10.1007/s11248-013-9727-2
- Uckun, F. M., Waddick, K. G., Mahajan, S., Jun, X., Takata, M., Bolen, J., et al. (1996). BTK as a mediator of radiation-induced apoptosis in DT-40 lymphoma B cells. *Science* 273, 1096–1100. doi: 10.1126/science.273.5278.1096
- Van De Lavoie, M.-C., Collarini, E. J., Leighton, P. A., Fesler, J., Lu, D. R., Harriman, W. D., et al. (2012). Interspecific germline transmission of cultured testicular cells. *PLoS ONE* 7:e35664. doi: 10.1371/journal.pone.0035664
- Van De Lavoie, M. C., Diamond, J. H., Leighton, P. A., Mather-Love, C., Heyer, B. S., Bradshaw, R., et al. (2006). Germline transmission of genetically modified primordial germ cells. *Nature* 441, 766–769. doi: 10.1038/nature04831
- Wang, K., Jin, Q., Ruan, D., Yang, Y., Liu, Q., Wu, H., et al. (2017). Cre-dependent Cas9-expressing pigs enable efficient *in vivo* genome editing. *Genome Res.* 27, 2061–2071. doi: 10.1101/gr.222521.117
- Wernery, U., Liu, C., Baskar, V., Guerinche, Z., Khazanehdari, K. A., Saleem, S., et al. (2010). Primordial germ cell-mediated chimera technology produces viable pure-line Houbara bustard offspring: potential for repopulating an endangered species. *PLoS ONE* 5:e15824. doi: 10.1371/journal.pone.0015824
- Whitworth, K. M., Rowland, R. R., Ewen, C. L., Tribble, B. R., Kerrigan, M. A., Cino-Ozuna, A. G., et al. (2015). Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nat. Biotechnol.* 34, 20–22. doi: 10.1038/nbt.3434
- Whyte, J., Glover, J. D., Woodcock, M., Brzeszczynska, J., Taylor, L., Sherman, A., et al. (2015). FGF, insulin, and SMAD signaling cooperate for avian primordial germ cell self-renewal. *Stem Cell Rep.* 5, 1171–1182. doi: 10.1016/j.stemcr.2015.10.008
- Wilson, N. H., and Stoeckli, E. T. (2012). In ovo electroporation of miRNA-based plasmids in the developing neural tube and assessment of phenotypes by DiI injection in open-book preparations. *J. Vis. Exp.* 16:4384. doi: 10.3791/4384
- Xu, W., Shao, Q., Zang, Y., Guo, Q., Zhang, Y., and Li, Z. (2015). Pigeon RIG-I function in innate immunity against H9N2 IAV and IBV. *Viruses* 7, 4131–4151. doi: 10.3390/v7072813
- Yasugi, S., and Nakamura, H. (2000). Gene transfer into chicken embryos as an effective system of analysis in developmental biology. *Dev. Growth Differ.* 42, 195–197. doi: 10.1046/j.1440-169x.2000.00500.x
- Yeh, S., Tsai, M.-Y., Xu, Q., Mu, X.-M., Lardy, H., Huang, K.-E., et al. (2002). Generation and characterization of androgen receptor knockout (ARKO) mice: an *in vivo* model for the study of androgen functions in selective tissues. *Proc. Natl. Acad. Sci. U.S.A.* 99, 13498–13503. doi: 10.1073/pnas.212474399
- Zazzaroni, F., Papa, S., De Smaele, E., and Franzoso, G. (2003). Cell signalling: cell survival and a Gadd45-factor deficiency: reply. *Nature* 424, 741–742. doi: 10.1038/424741b
- Zhang, X., Yan, Z., Li, X., Lin, W., Dai, Z., Yan, Y., et al. (2016). GADD45 β , an anti-tumor gene, inhibits avian leukosis virus subgroup J replication in chickens. *Oncotarget* 7, 68883–68893. doi: 10.18632/oncotarget.12027
- Zhu, L., Van De Lavoie, M.-C., Albanese, J., Beenhouwer, D. O., Cardarelli, P. M., Cuisson, S., et al. (2005). Production of human monoclonal antibody in eggs of chimeric chickens. *Nat. Biotechnol.* 23, 1159–1169. doi: 10.1038/nbt1132

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Sid and Schusser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Including Phenotypic Causal Networks in Genome-Wide Association Studies Using Mixed Effects Structural Equation Models

Mehdi Momen¹, Ahmad Ayatollahi Mehrgardi^{1*}, Mahmoud Amiri Roudbar¹, Andreas Kranis², Renan Mercuri Pinto^{3,4}, Bruno D. Valente⁴, Gota Morota⁵, Guilherme J. M. Rosa^{4,6} and Daniel Gianola^{4,6,7}

¹ Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of Kerman, Kerman, Iran, ² Roslin Institute, University of Edinburgh, Midlothian, United Kingdom, ³ Department of Exact Sciences, University of São Paulo-Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, Brazil, ⁴ Department of Animal Sciences, University of Wisconsin, Madison, WI, United States, ⁵ Department of Animal and Poultry Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, ⁶ Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, United States, ⁷ Department of Dairy Science, University of Wisconsin, Madison, WI, United States

OPEN ACCESS

Edited by:

John Anthony Hammond,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Fabyano Fonseca Silva,
Universidade Federal de Viçosa, Brazil
Gregor Gorjanc,
University of Edinburgh,
United Kingdom

*Correspondence:

Ahmad Ayatollahi Mehrgardi
mehrgardi@uk.ac.ir

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 21 June 2018

Accepted: 18 September 2018

Published: 09 October 2018

Citation:

Momen M, Ayatollahi Mehrgardi A, Amiri Roudbar M, Kranis A, Mercuri Pinto R, Valente BD, Morota G, Rosa GJM and Gianola D (2018) Including Phenotypic Causal Networks in Genome-Wide Association Studies Using Mixed Effects Structural Equation Models. *Front. Genet.* 9:455. doi: 10.3389/fgene.2018.00455

Network based statistical models accounting for putative causal relationships among multiple phenotypes can be used to infer single-nucleotide polymorphism (SNP) effect which transmitting through a given causal path in genome-wide association studies (GWAS). In GWAS with multiple phenotypes, reconstructing underlying causal structures among traits and SNPs using a single statistical framework is essential for understanding the entirety of genotype-phenotype maps. A structural equation model (SEM) can be used for such purposes. We applied SEM to GWAS (SEM-GWAS) in chickens, taking into account putative causal relationships among breast meat (BM), body weight (BW), hen-house production (HHP), and SNPs. We assessed the performance of SEM-GWAS by comparing the model results with those obtained from traditional multi-trait association analyses (MTM-GWAS). Three different putative causal path diagrams were inferred from highest posterior density (HPD) intervals of 0.75, 0.85, and 0.95 using the inductive causation algorithm. A positive path coefficient was estimated for BM→ BW, and negative values were obtained for BM→ HHP and BW→ HHP in all implemented scenarios. Further, the application of SEM-GWAS enabled the decomposition of SNP effects into direct, indirect, and total effects, identifying whether a SNP effect is acting directly or indirectly on a given trait. In contrast, MTM-GWAS only captured overall genetic effects on traits, which is equivalent to combining the direct and indirect SNP effects from SEM-GWAS. Although MTM-GWAS and SEM-GWAS use the similar probabilistic models, we provide evidence that SEM-GWAS captures complex relationships in terms of causal meaning and mediation and delivers a more comprehensive understanding of SNP effects compared to MTM-GWAS. Our results showed that SEM-GWAS provides important insight regarding the mechanism by which identified SNPs control traits by partitioning them into direct, indirect, and total SNP effects.

Keywords: causal structure, GWAS, multiple traits, path analysis, SEM, SNP effect

INTRODUCTION

Genome-wide association studies (GWAS) have become a standard approach for investigating relationships between common genetic variants in the genome (e.g., single-nucleotide polymorphisms, SNPs) and phenotypes of interest in human, plant, and animal genetics (Hayes and Goddard, 2010; Brachi et al., 2011; Wang et al., 2012). A typical GWAS is based on univariate linear or logistic regression of phenotypes on genotypes for each SNP individually while often adjusting for the presence of nuisance covariates (Hayes and Goddard, 2010; Sikorska et al., 2013). A statistically significant association indicates that SNPs may be in strong linkage disequilibrium (LD) with quantitative trait loci (QTL) that contribute to the trait etiology. Alternatively, multi-trait model GWAS (MTM-GWAS) can be used to test for genetic associations among a set of traits (Korte et al., 2012; O'Reilly et al., 2012; Zhou and Stephens, 2012). It has been established that MTM-GWAS reduces false positives and increases the statistical power of association tests, explaining the recent popularity of this method. MTM-GWAS can be used to study genetic associations among a set of traits. However, it does not consider various cryptic biological signals that may affect a trait of interest, either directly or indirectly through other intermediate traits.

Complex traits are the product of various cryptic biological signals that may affect a trait of interest either directly or indirectly through other intermediate traits (Falconer and Mackay, 1996). A standard regression cannot describe such complex relationships between traits and QTLs properly. For instance, some traits may simultaneously act as both dependent and independent variables. Structural equation modeling (SEM) is an extended version of Wright's path analysis (Wright, 1921; Gianola and Sorensen, 2004) that offers a powerful technique for modeling causal networks. In a complex genotype-phenotype setting involving many traits, a given trait can be influenced not only by genetic and systematic factors but also by other traits (as covariates). Here, QTLs may not affect the target trait directly; instead, the effects may be mediated by upstream traits in a causal network. Indirect effects may therefore constitute a proportion of perceived pleiotropy, and these concepts apply to sets of heritable traits, organized as networks, that are common in biological systems. An example from dairy cattle production systems, described by Gianola and Sorensen (2004), is that higher milk yield increases the risk of a particular disease, such as mastitis, while the prevalence of the disease may negatively affect milk yield. As another example, Varona et al. (2007) explored a causal link from litter size to average piglet weight in two pig breeds. In humans, obesity is a key factor influencing insulin resistance, which subsequently causes type 2 diabetes. Lists of causal networks across human diseases and candidate genes are described in Kumar and Agrawal (2013) and Schadt (2016).

Although MTM-GWAS is a valuable approach, it only captures correlations or associations among traits and does not provide information about causal relationships. Knowledge of the causal structures underlying complex traits is essential, as correlation does not imply causation. For example, a correlation between two traits, T1 and T2, could be attributed to a direct

effect of T1 on T2 or T2 on T1, or to additional variables that jointly influence both traits (Rosa et al., 2011). Likewise, if we know a "causal" SNP is linked to a QTL, we can imagine three possible scenarios with respect to T1: (1) causal ($SNP \rightarrow T1 \rightarrow T2$), (2) reactive ($SNP \rightarrow T2 \rightarrow T1$), or (3) independent ($T1 \leftarrow SNP \rightarrow T2$). Scenarios (1) and (2) do not cause pleiotropy but produce association.

A SEM methodology has the ability to handle complex genotype-phenotype maps in GWAS, placing an emphasis on causal networks (Li et al., 2006). Therefore, SEM-based GWAS (SEM-GWAS) may provide a better understanding of biological mechanisms and of relationships among a set of traits than MTM-GWAS. SEM can potentially decompose the total SNP effect on a trait into direct and indirect (i.e., mediated) contributions. However, SEM-derived GWAS has yet not been discussed or applied fully in quantitative genetic studies yet. Our objective was to illustrate the potential utility of SEM-GWAS by using three production traits in broiler chickens genotyped for a battery of SNPs as a case example.

MATERIALS AND METHODS

Data Set

The analysis included records for 1,351 broiler chickens provided by Aviagen Ltd. (Newbridge, Scotland) for three phenotypic traits: ultrasound of breast muscle (BM) at 35 days of age, body weight (BW), and hen-house egg production (HHP), defined as the total number of eggs laid between weeks 28 and 54 per bird. The sample consisted of 274 full-sib families, 326 sires, and 592 dams. More details regarding population and family structure were provided by Momen et al. (2017). A pre-correction procedure was performed on the phenotypes to account for systematic effects such as sex, hatch week, pen, and contemporary group for BM and BW. HHP was corrected for random hatch effects, with a general mean as the sole fixed effect.

Each bird was genotyped for 580,954 SNP markers with a 600k Affymetrix SNP (Kranis et al., 2013) chip (Affymetrix, Inc., Santa Clara, CA, USA). The Beagle software program (Browning and Browning, 2007) was used to impute missing SNP genotypes, and quality control was performed using PLINK version 1.9 (Purcell et al., 2007). Markers with minor allele frequencies (MAF) < 1%, call rate < 95%, and Hardy-Weinberg equilibrium (Chi-square test p -value threshold was 10^{-6}) were removed. The main reason for conducting the HWE test was to remove SNPs with potential genotyping error. Finally, 354,364 autosomal SNP markers were included in the analysis.

Multiple-Trait Model for GWAS

MTM-GWAS is a single-trait GWAS model extended to multi-dimensional responses. When only considering additive effects of SNPs, the phenotype of a quantitative trait using the single-trait model can be described as:

$$y_i = w_{ij}s_j + e_i \quad (1)$$

where y_i is the phenotypic trait of individual i , $w_j = (w_1, \dots, w_p)$ is the number of A alleles (i.e., $w_j \in \{0, 1, 2\}$) in the genotype

of SNP marker j , and s_j is the allele substitution effect for SNP marker j . Strong LD between markers and QTLs coupled with an adequate marker density increases the chance of detecting marker and phenotype associations. Hypothesis testing is typically used to evaluate the strength of the evidence of a putative association. Typically, a t -test is applied to obtain p -values, and the statistic is $T_{ij} = \frac{\hat{s}_j}{se(\hat{s}_j)}$, where \hat{s} is the point estimate of the j -th SNP effect and $se(\hat{s}_j)$ is its standard error.

The single locus model described above is naive for a complex trait because the data typically contain hidden population structure and individuals have varying degrees of genetic similarity (Listgarten et al., 2012; Gianola et al., 2016). Therefore, accounting for covariance structure induced by genetic similarity is expected to produce better inferences (Kennedy et al., 1992). Ignoring effects that reveal genetic relatedness inflates the residual terms and compromises the ability to detect association. A random effect g_i , including a covariance matrix reflecting pairwise similarities between additive genetic effects of individuals, can be included to control population stratification. The similarity metrics can be derived from pedigree information or from whole-genome marker genotypes. This model, extended for analysis of t traits, is given by:

$$Y = Ws + g + \varepsilon \quad (2)$$

where Y is the pre-adjusted phenotypic value measured on each birds, W as previously defined, represent the incidence matrix of genotype codes, s is the vector of additive marker effect, g is the vector of random polygenic effect, $g \sim N(0, \sum_g \otimes K)$, and ε represents the residual vector, $\varepsilon \sim N(0, \sum_\varepsilon \otimes I)$. Here \otimes denotes the Kronecker product. The covariance matrices were:

$$\sum_g = \begin{bmatrix} \sigma_{g(BM)}^2 & \sigma_{g(BM,BW)} & \sigma_{g(BM,HHP)} \\ \sigma_{g(BW)}^2 & \sigma_{g(BW,HHP)} & \sigma_{g(HHP)}^2 \\ \text{Symmetric} & & \end{bmatrix} \text{ and}$$

$$\sum_\varepsilon = \begin{bmatrix} \sigma_{\varepsilon(BM)}^2 & \sigma_{\varepsilon(BM,BW)} & \sigma_{\varepsilon(BM,HHP)} \\ \sigma_{\varepsilon(BW)}^2 & \sigma_{\varepsilon(BW,HHP)} & \sigma_{\varepsilon(HHP)}^2 \\ \text{Symmetric} & & \end{bmatrix}.$$

The positive definite matrix K may be a genomic relationship matrix (G) computed from marker data, or a pedigree-based matrix (A) computed from genealogical information. The A matrix describes the expected additive similarity among individuals, while G measures the realized fraction of alleles shared. Genomic relationship matrices can be derived in several ways (VanRaden, 2008; Yang et al., 2010; Forni et al., 2011). Here, we used the form proposed by VanRaden (2008):

$$G = \frac{MM'}{2 \sum p_j q_j} \quad (3)$$

where M is an $n \times p$ matrix of centered SNP genotypes and p_j and $q_j = 1 - p_j$ are the allele frequencies at marker locus j . We evaluated both A and G in the present study.

Structural Equation Model Association Analysis

A SEM consists of two essential parts: a measurement model and a structural model. The measurement model depicts the connections between observable variables and their corresponding latent variables (Anderson and Gerbing, 1988). The measurement model is also known as confirmatory factor analysis. The critical part of a SEM is the structural model, which can have three forms (Raykov and Marcoulides, 2012). The first consists of observable exogenous and endogenous variables. This model is a restricted version of a SEM known as path analysis (Wright, 1921). The second form explains the relationship between exogenous and endogenous variables that are only latent. The third type is a model consisting of both manifest and latent variables.

SEM can be applied to GWAS as an alternative to MTM-GWAS to study how different causal paths mediate SNP effects on each trait. The following SEM model was considered:

$$Y = \Lambda Y + Ws + g + \varepsilon \quad (4)$$

where Λ is a $t \times t$ matrix of regression coefficients or structural coefficients (typically lower-triangular) according to the learned causal structure from the residuals and the diagonal matrix filled with zeros:

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 \\ \lambda_{(BM \rightarrow BW)} & 0 & 0 \\ \lambda_{(BM \rightarrow HHP)} & \lambda_{(BW \rightarrow HHP)} & 0 \end{bmatrix}$$

The vectors g and ε are assumed to have a joint distribution $\begin{bmatrix} g \\ \varepsilon \end{bmatrix} = N \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sum_g \otimes K & 0 \\ 0 & \Psi \end{bmatrix} \right\}$, and the residual covariance

matrix is a diagonal as $\Psi = \begin{bmatrix} \sigma_{\varepsilon(BM)}^2 & 0 & 0 \\ 0 & \sigma_{\varepsilon(BW)}^2 & 0 \\ 0 & 0 & \sigma_{\varepsilon(HHP)}^2 \end{bmatrix}$. The

remaining terms are as presented earlier with one important difference: the SNP effects are not interpreted as overall effects on trait t but instead represent direct effects on trait t . Additional indirect effects from the same SNP may be mediated by phenotypic traits in C . Each marker is entered into Equation (4) separately, and its significance is tested. For a discussion of how SEM represents genetic signals on each trait through multiple causal paths, see Wu et al. (2010) and Jamrozik and Schaeffer (2011). Despite the difference in interpretation, the distribution of the vector of polygenic effects is assumed to be the same as in the MTM-GWAS model. The same applies to residual terms within a trait. We also consider trait-specific residuals to be independent within an individual. This restriction is required to render structural coefficients likelihood-identifiable. In addition, the interpretation of inferences as having a causal meaning requires imposing the restriction that the residuals' joint distribution be interpreted as the causal sufficiency assumption (Pearl, 2009). In the present study, all exogenous and endogenous variables were observable, and there was no latent variable. Hence, causal structure was assumed between the endogenous variables BM, BW, and HHP.

We considered the following GWAS models with their causal structures were recovered by the inductive causation (IC) algorithm (Pearl, 2009): (1) MTM-GWAS with pedigree-based kinship **A** (MTM-A) or marker-based kinship **G** (MTM-G), and (2) SEM-GWAS with **A** (SEM-A) or **G** (SEM-G). Although nuisance covariates such as environmental factors can be omitted in the graph, they may be incorporated into the models as exogenous variables. The SEM representation allowed us to decompose SNP effects into direct, indirect, and total effects.

A direct SNP effect is the path coefficient between a SNP as an exogenous variable and a dependent variable without any causal mediation by any other variable. The indirect effects of a SNP are those mediated by at least one other intervening endogenous variable. Indirect effects are calculated by multiplying path coefficients for each path linking the SNP to an associated variable, and then summing over all such paths (Mi et al., 2010a; Jiang et al., 2013). The overall effect is the sum of all direct and indirect effects. By explicitly accounting for complex relationship structure among traits in such a way, SEM provides a better understanding of a genome-wide SNP analysis by allowing us to decompose effects into direct, indirect, and overall effects within a predefined casual framework (Nock and Zhang, 2011). MTM-GWAS and SEM-GWAS were compared with the logarithm of the likelihood function (log L), Akaike's Information Criterion (AIC), and the Bayesian Information Criterion (BIC). The model providing the lowest values for these information criteria is considered to fit the data better. MTM-GWAS and SEM-GWAS were fitted using the SNP Snappy strategy (Meyer and Tier, 2012), which is implemented in the Wombat software program (Meyer, 2007). The outputs were a vector of multiple SNP effect estimates, $\hat{s} = [\hat{s}_{BM}, \hat{s}_{BW}, \hat{s}_{HHP}]$, with corresponding standard errors and respective t-values.

Searching for a Phenotypic Causal Network in a Mixed Model

In the SEM-GWAS formulation described earlier, the structure of the underlying causal phenotypic network needs to be known. Because this is not so in practice, we used a causal inference algorithm to infer the structure. Residuals are assumed to be independent in all SEM analyses, so associations between observed traits are viewed as due to causal links between traits and by correlations among genetic values (i.e., g_1 , g_2 , and g_3). Thus, to eliminate confounding problems when inferring the underlying network among traits, we used the approach of Valente et al. (2010) to search for acyclic causal structures through conditional independencies on the distribution of the phenotypes, given the genetic effects. A causal phenotypic network was inferred in two stages: (1) an MTM model (Henderson and Quaas, 1976) was employed to estimate covariance matrices of additive genetic effects and of residuals, and (2) the causal structure among phenotypes from the covariance matrix between traits, conditionally on additive genetic effects, was inferred by the IC algorithm. The residual (co)variance matrix was inferred using Bayesian Markov-chain Monte Carlo (Valente et al., 2010; Wu et al., 2010), with

samples drawn from the posterior distribution. The reason for our use of the residual (co)covariances is that the residual structure could bear information from the joint distribution of all phenotypic traits conditional on their polygenic effects, such that they correct the confounding issues caused by such effects when the traits are genetically correlated (Pearl, 2009). For each query testing statistical independence between traits y_t and $y_{t'}$, the posterior distribution of the residual partial correlation $\rho_{y_t, y_{t'}} | h$ was obtained, where h is a set of variables (traits) that are independent. Three highest posterior density (HPD) intervals of 0.75, 0.85, and 0.95 were used to make statistical decisions for SEM-GWAS. We thus considered SEM-A75 (HPD > 0.75), SEM-A85 (HPD > 0.85), SEM-A95 (HPD > 0.95), and SEM-G75 (HPD > 0.75). An HPD interval that does not contain zero declares y_t and $y_{t'}$ to be conditionally dependent.

RESULTS

Figure 1 shows phenotypic relationship structures recovered by the IC algorithm for the three different HPD intervals. Edges connecting two traits represent non-null partial correlations as indicated by HPD intervals. We compared the two MTM-GWAS and four SEM-GWAS by using the three chicken traits (BW, BM, and HHP). Fully recursive (there is at least one incoming OR outgoing edge for each node) SEM-A75 and SEM-G75 graphs revealed direct effects of BM on BW and HHP, and those of BW on HHP, as well as an indirect effect of BM on HHP. In addition, SEM-A85 detected a direct effect of BM on BW, the direct effect of BW on HHP, and the indirect effect of BM on HHP mediated by BW. Finally, SEM-A95 only identified a direct effect of BM on BW because of a statistically stringent HPD cutoff imposed. SEM-G85 and SEM-G95 were not explored further because they produced the same results as SEM-A85 and SEM-A95.

Given the causal structures inferred from the IC algorithm, the following SEM was fitted:

$$\begin{cases} y_1 = \mu + Z_i g_1 + W_{ij} s_{jl} + \varepsilon_i \\ y_2 = \mu + \lambda_{21} y_1 + Z_i g_2 + W_{ij} s_{jl} + \varepsilon_i \\ y_3 = \mu + \lambda_{31} y_1 + \lambda_{32} y_2 + Z_i g_3 + W_{ij} s_{jl} + \varepsilon_i \end{cases} \quad (5)$$

Note that only a small number of the entries in the structural coefficient matrix (λ in Equation 5) are non-zero due to sparsity. These non-zero entries specify the effect of one phenotype on other phenotypes. The corresponding directed acyclic graph is shown in **Figure 2** assuming the causal relationships among the three traits, where y_1 , y_2 , and y_3 represent BM, BW, and HHP, respectively; SNP_j is the genotype of the j -th SNP; s_{jl} is the direct SNP effect on trait l ; and the remaining variables are as presented earlier. This diagram depicts a fully recursive structure in which all recursive relationships among the three phenotypic traits are shown. Arrows represent causal connections, whereas double-headed arrows between polygenic effects are correlations.

We examined the fit of each model implemented to assess how well it describes the data (**Table 1**). Varona et al. (2007) and recently Valente et al. (2013) showed that re-parametrization

and reduction of a SEM mixed model yield the same joint probability distribution of observation as in MTM, suggesting that the expected likelihood of SEM and MTM should be similar. As expected, SEM-GWAS and MTM-GWAS showed very similar results (e.g., SEM-A75 vs. MTM-A and SEM-G75 vs. MTM-G). Among the models considered, those involving **G** exhibited slightly better fits. SEM-A85 and SEM-A95, sharing a subset of the SEM-A75 structure, presented almost identical AIC and BIC values. Since these results imply that the recursive model and standard mixed model for GWAS are statistically equivalent in terms of the fitting criteria, the focus of the remainder of the analysis will be on the modeling of SNP (or QTL) effects in the SEM context (SEM-A75 or SEM-G75) as an extension of MTM, which accounts for recursive links among the three measured traits.

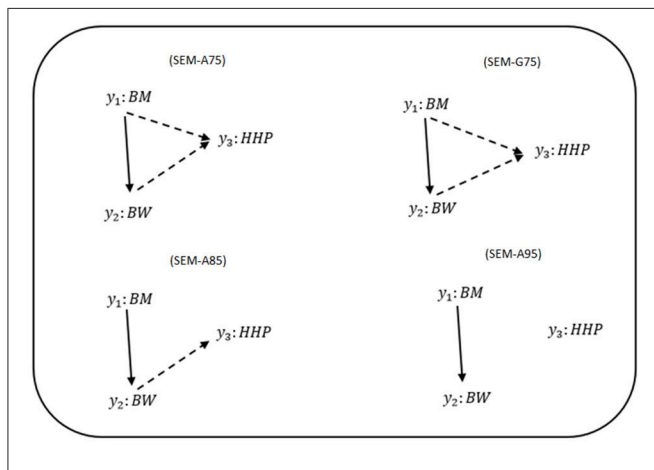


FIGURE 1 | Causal graphs inferred using the IC algorithm among three traits: breast meat (BM), body weight (BW), and hen-house production (HHP) in the chicken data. SEM-A75 and SEM-G75 were the inferred fully recursive causal structures with HPD > 0.75 and corrected for genetic confounder using A (pedigree-based) and G (marker-based) matrices. SEM-A85 and SEM-A95 were obtained with HPD > 0.85 and HPD > 0.95, respectively, corrected with A. Arrows indicate direction of causal relationships. Dashed lines indicate negative coefficients, and the continuous arrows indicate positive coefficients.

Structural Coefficients

Table 2 presents the causal structural path coefficients for endogenous variables (BM, BW, and HHP). All models have positive effects for $BM \rightarrow BW$, whereas the $BM \rightarrow HHP$ and $BW \rightarrow HHP$ relationships have negative path coefficients. The latter confirmed the fact that chicken breeding is divided into broiler and layer sections due to the negative genetic correlation between BW and HHP.

Also shown in **Table 2** are the magnitudes of the SEM structural coefficient reflecting the intensity of the causality. The positive coefficient λ_{21} quantifies the (direct) causal effect of BM on BW. This suggests that a 1-unit increase in BM results in a λ_{21} -unit increase in BW. Likewise, the negative causal effects λ_{31} and λ_{32} offer the same interpretation.

Decomposition of SNP Effect Paths Using a Fully Recursive Model

We can decompose SNP effects into direct and indirect effects using **Figure 2**. The direct effect of the SNP j on y_3 (HHP) is given by $d_{SNP_j \rightarrow y_3} : \hat{S}_{j(y_3)}$, where d denotes the direct effect. Note there are only one direct and many indirect paths. We find three indirect paths from SNP_j to y_3 mediated by y_1 and y_2 (i.e., the nodes formed by other traits). The first indirect effect is $ind_{(1)SNP_j \rightarrow y_3} : \lambda_{32}(\lambda_{21}\hat{S}_{j(y_1)})$ in the path mediated by y_1 and y_2 , where ind denotes the indirect effect. The second indirect effect $ind_{(2)SNP_j \rightarrow y_3} : \lambda_{32}\hat{S}_{j(y_2)}$, is mediated by y_2 . The last indirect effect, is $ind_{(3)SNP_j \rightarrow y_3} : \lambda_{31}\hat{S}_{j(y_1)}$, mediated by y_1 . Therefore, the overall effect is given by summing all four paths, $T_{SNP_j \rightarrow y_3} : \lambda_{32}(\lambda_{21}\hat{S}_{j(y_1)}) + \lambda_{32}\hat{S}_{j(y_2)} + \lambda_{31}\hat{S}_{j(y_1)} + \hat{S}_{j(y_3)}$. The fully recursive model of the overall SNP effect is then:

$$\begin{cases} T_{\hat{S}_j \rightarrow y_1} : \hat{S}_{j(y_1)} \\ T_{\hat{S}_j \rightarrow y_2} : \lambda_{21}(\hat{S}_{j(y_1)}) + \hat{S}_{j(y_2)} \\ T_{\hat{S}_j \rightarrow y_3} : \lambda_{32}[\lambda_{21}(\hat{S}_{j(y_1)}) + \hat{S}_{j(y_2)}] + \lambda_{31}(\hat{S}_{j(y_1)}) + \hat{S}_{j(y_3)} \end{cases} \quad (6)$$

For y_1 (BM), there is only one effect, so the overall effect is equal to the direct effect. For y_2 (BW) and y_3 (HHP), direct and indirect SNP effects are involved. There are two paths for

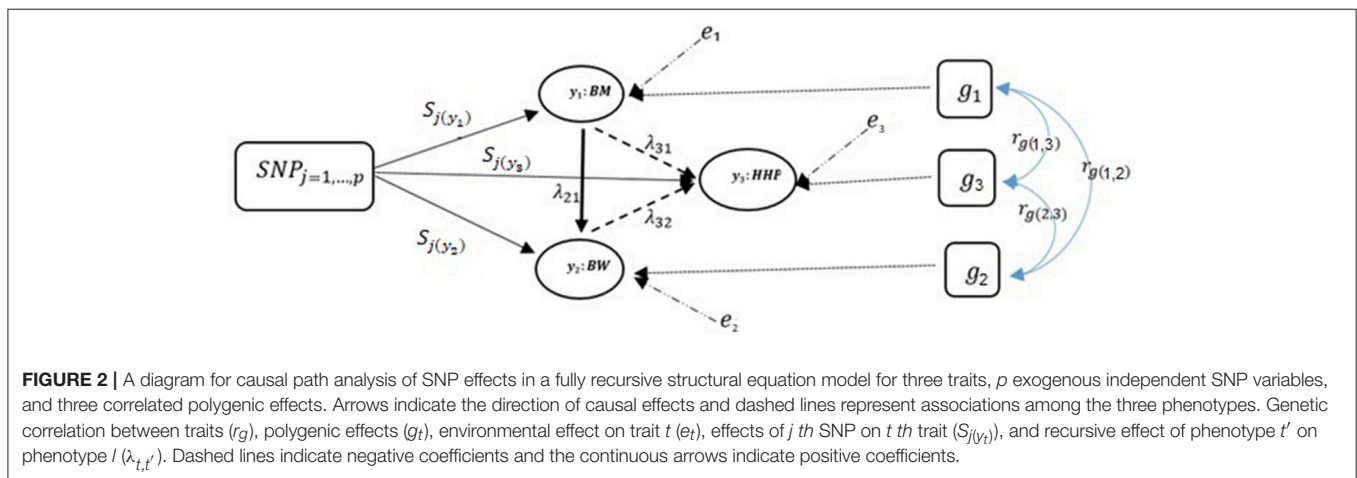


FIGURE 2 | A diagram for causal path analysis of SNP effects in a fully recursive structural equation model for three traits, p exogenous independent SNP variables, and three correlated polygenic effects. Arrows indicate the direction of causal effects and dashed lines represent associations among the three phenotypes. Genetic correlation between traits (r_g), polygenic effects (g_t), environmental effect on trait t (e_t), effects of j th SNP on t th trait ($S_{j(y_t)}$), and recursive effect of phenotype t' on phenotype t ($\lambda_{t',t}$). Dashed lines indicate negative coefficients and the continuous arrows indicate positive coefficients.

TABLE 1 | Model comparison criteria: logarithm of the restricted maximum likelihood function (log L), Akaike's information criteria (AIC), Schwarz Bayesian information criteria (BIC) were used evaluate model fit for two multiple trait models (MTM) and four structural equation models (SEM).

| Model | Maximum log L | −1/2 AIC | −1/2 BIC |
|---------|---------------|-----------|-----------|
| MTM-A | −7093.480 | −7105.48 | −7142.436 |
| SEM-A75 | −7098.370 | −7110.415 | −7147.321 |
| SEM-A85 | −7095.188 | −7107.188 | −7144.143 |
| SEM-A95 | −7097.517 | −7109.517 | −7146.470 |
| MTM-G | −6529.270 | −6541.276 | −6578.232 |
| SEM-G75 | −6537.391 | −6549.391 | −6586.34 |

A, pedigree-based relationship matrix, **G**, VanRaden's marker-based relationship matrix. MTM-A and MTM-G denote MTM-GWAS models coupled with the **A** and **G** matrices, respectively. SEM-A75, SEM-A85 and SEM-A95 represent SEM-GWAS models with HPD > 75, 85, and 95% values with the **A** matrix, respectively. SEM-G75 is a SEM-GWAS model with HPD > 75 coupled with the **G** matrix.

TABLE 2 | Estimates of three causal structural coefficients (λ) derived from four different structural models.

| Path | Structural models | | | |
|--|-------------------|---------|---------|---------|
| | SEM-A75 | SEM-G75 | SEM-A85 | SEM-A95 |
| $\lambda_{BM \rightarrow BW}(\lambda_{21})$ | 2.13 | 2.19 | 2.14 | 2.14 |
| $\lambda_{BM \rightarrow HHP}(\lambda_{31})$ | −0.17 | −0.280 | *** | *** |
| $\lambda_{BW \rightarrow HHP}(\lambda_{32})$ | −0.27 | −0.096 | −0.31 | *** |

BM, breast meat; BW, body weight; HHP, hen-house production. SEM-75: HPD > 0.75. SEM-G75: HPD > 0.75. SEM-A85: HPD > 0.85. SEM-A95: HPD > 0.95. ***Represents path coefficient was not estimated because there was no corresponding path in the inferred structure.

y_2 : one indirect, $ind_{S_j \rightarrow y_2} : \hat{S}_{j(y_1)} \rightarrow y_1 \rightarrow y_2$, and one direct, $d_{S_j \rightarrow y_2} : \hat{S}_{j(y_2)} \rightarrow y_2$. Here, the SNP effect is direct and mediated thorough other phenotypes according to causal networks in SEM-GWAS (Figures 1, 2). For instance, the overall SNP effect for y_3 into four direct and indirect paths is $T_{\hat{S}_j \rightarrow y_3} : \lambda_{32}\lambda_{21}\hat{S}_{j(y_1)} + \lambda_{32}\hat{S}_{j(y_1)} + \lambda_{31}\hat{S}_{j(y_1)} + \hat{S}_{j(y_3)}$.

The scatter plots in Figure 3 compare the estimated total effects for HHP ($T_{\hat{S}_j \rightarrow y_3}$) obtained from SEM-GWAS and those from MTM-GWAS. We observed good agreement between SEM-GWAS and MTM-GWAS. The total SNP signals derived from SEM and MTM are the same but SEM provides biologically relevant additional information.

Figures S1–S4 present scatter plots of MTM-GWAS and SEM-GWAS signals (SEM-A75, SEM-G75, SEM-A85, and SEM-A95) for the $BM \rightarrow BW$ path, which was a common path across all SEM-GWAS considered. These two traits have a genetic correlation of 0.5 (results not shown). We partitioned the SEM causal link into direct, indirect, and overall effects based on directed links inferred from the IC algorithm with HPD > 0.85, whereas MTM-GWAS captures an overall SNP effect on BW. Scatter plots of the overall effects from SEM-GWAS and those of the total effects from MTM-GWAS indicated almost perfect agreement (top left plots, Figures S1–S4). We also observed

concomitance between estimated overall and direct effects (top right plots, Figures S1–S4). In contrast, there was less agreement in the magnitude of the SNP effects when comparing overall vs. indirect effects (bottom left plots, Figures S1–S4). There was no linear relationship between the indirect and direct SNP effects (bottom right plots, Figures S1–S4). In short, genetic signals detected in SEM-GWAS were close to those of MTM-GWAS for overall effects because both models are based on a multivariate approach with the same covariance matrix. In all SEM-GWAS, results showed that direct effects contributed to overall effects more than the indirect effects.

Manhattan Plot of Direct, Indirect, and Overall SNP Effects

Figure 4 depicts a Manhattan plot summarizing the magnitude of direct (SEM-75A), indirect (SEM-75A), and overall SNP effects (MTM-75A). We plotted the decomposed SNP effects on BW along chromosomes to visualize estimated marker effects from SEM-GWAS and MTM-GWAS. The indirect and direct effects provide a view of SNP effects from a perspective that is not available for the total effect of MTM-GWAS. For instance, there were two estimated SNP effects on chromosomes 1 and 2 that deserve particular attention. These two SNPs are highlighted with black circles and red ovals. The overall effect of the first SNP consisted of large indirect and small direct effects on BM, whereas the opposite pattern was observed for the second SNP, which showed large direct and small indirect effects. Although the overall effects of these SNPs were similar (top Manhattan plot, Figure 4), use of decomposition allowed us to determine that the trait of interest is affected in different manners: the second SNP effect acted directly on BW without any mediation by BM, whereas the first SNP reflected a large effect mediated by BM on BW. Collectively, new insight regarding the direction of SNP effects can be obtained using the SEM-GWAS methodology.

The corresponding Manhattan plot based on $-\log_{10}(p\text{-values})$ is shown in Figure S5. As with the magnitude of effect sizes, the results showed that $-\log_{10}(p\text{-values})$ of estimated overall effects from SEM-A75 and those from MTM-A75 yielded the same significant peaks. We found that some significant indirect SNP effects reached genome-wide significance after correction for multiple-testing using a 5% FDR threshold level (2.752). The most significant SNPs were on chromosomes 1 and 4 (GGA1 and GGA4).

As an illustration, the six most significant SNPs with the highest $-\log_{10}(p\text{-values})$ for each type of decomposed SNP effect are presented in Table 3. Seven candidate genes were identified near the significant SNPs derived from the SNP effects decomposition, with two on GGA7 (OLA1 and ZNF385B), one on GGA3 (EPHA7), three on GGA4 (LOC422264, LOC422265, and MAEA), and one on GGA14 (GRIN2A). We found that only genes on GGA4 and GGA1 are linked to significant indirect SNP effects that impact HHP. Some studies reported QTLs for BM on GGA1 and for BW on GGA4, stating that these genomic regions contain QTLs related to abdominal fat and growth traits that were detected across

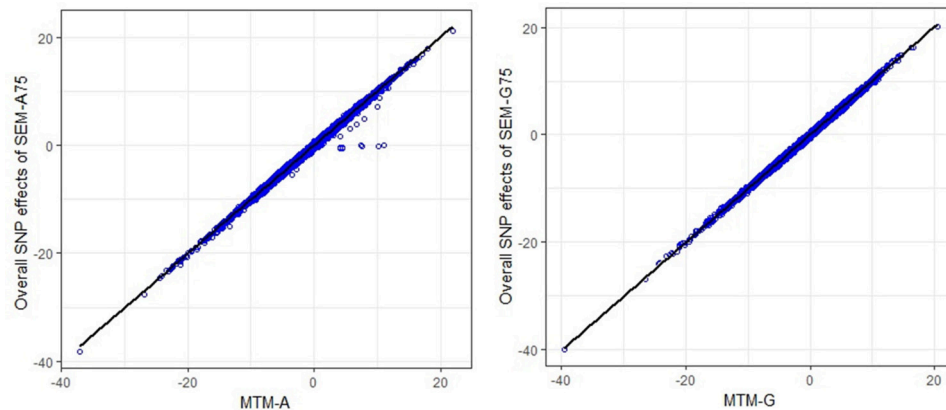


FIGURE 3 | Comparison of multiple trait (MTM) and fully recursive overall SNP effects obtained with **A** (pedigree-based) and **G** (marker-based) from structural equation modeling (SEM)-based GWAS. Overall effects in SEM are the sum of all direct and indirect effects. HHP, hen-house egg production.

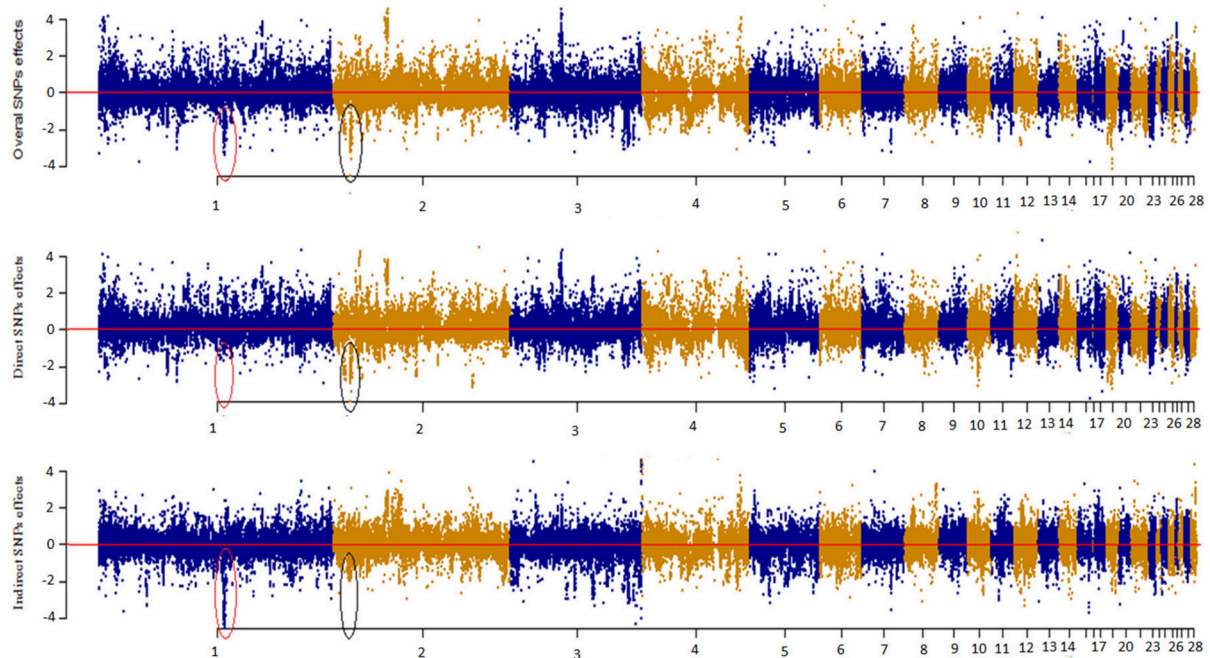


FIGURE 4 | Manhattan plot showing overall, direct and indirect SNP effects using a full recursive model based on **A** matrix for body weight (BW).

diverse chicken populations (Sun et al., 2013; Van Goor et al., 2015). One of the two detected genes on GGA14, i.e., *GRIN2A*, which was linked to the SNP Gga_rs313620413, showed significant direct and overall SNPs effects using SEM as well as MTM. Collectively, Gga_rs15390496, Gga_rs16591372, and Gga_rs313620413 SNPs on GGA3, GGA7, and GGA14, which were linked to *EPHA7*, *OLA1*, and *GRIN2A*, respectively, represent candidate genes identified from overall effects of both SEM and MTM (Table 3).

We noted that the six SNPs selected according to the $-\log_{10}$ (p -values) from the direct effect on HHP (i.e., $d_{SNP_j \rightarrow y_{(HHP)}}$)

had small indirect effects ranging from -0.9018 to 0.2983 . These indirect effects were negligible compared with their corresponding direct and total effects. Also, exploring the indirect effect sizes of the six most significant SNPs showed that indirect effects that are transmitted through inferred causal networks have the ability to change the magnitude of overall SNP effects, even changing them to the opposite direction (i.e., from positive to negative or vice versa).

It should also be noted that the estimated additive SNP effects obtained from the four SEM-GWAS can be used for inferring pleiotropy. For instance, a pleiotropic QTL may have a large

TABLE 3 | Six most significant SNPs selected according $-\log_{10}$ (p -values) and their effects, using the full recursive SEM (SEM-A75) and MTM (MTM-A75).

| | CHR | SNP name | Candidate genes | $-\log_{10}$ (p -values) | | | | Type of SNP effect | | | |
|-------------------------------|-----|-----------------|-----------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|
| | | | | $ds_{j \rightarrow y(HHP)}$ | $inds_{j \rightarrow y(HHP)}$ | $TS_{j \rightarrow y(HHP)}$ | $MTMS_{j \rightarrow y(HHP)}$ | $ds_{j \rightarrow y(HHP)}$ | $inds_{j \rightarrow y(HHP)}$ | $TS_{j \rightarrow y(HHP)}$ | $MTMS_{j \rightarrow y(HHP)}$ |
| Top SNPs for direct effects | 14 | Gga_rs313620413 | GRIN2A | 7.4242 | 0.1499 | 9.6599 | 7.4525 | -5.7827 | -0.0498 | -5.8326 | -5.78511 |
| | 7 | Gga_rs16591372 | OLA1 | 7.0868 | 0.2220 | 9.0119 | 6.9783 | -22.5681 | 0.2983 | -22.2698 | -22.3520 |
| | 3 | Gga_rs15390496 | EPHA7 | 7.0209 | 0.2214 | 8.6122 | 7.0297 | -22.4233 | -0.2149 | -22.6382 | -22.4098 |
| | 1 | Gga_rs314001234 | – | 7.0147 | 1.1067 | 9.0710 | 7.1653 | -26.6538 | -0.9018 | -27.5556 | -26.9360 |
| | 7 | Gga_rs315626061 | – | 6.8300 | 0.3360 | 8.9974 | 6.9529 | 5.1767 | 0.0910 | 5.26783 | 5.22295 |
| | 7 | Gga_rs316509306 | – | 6.8241 | 0.3442 | 8.9952 | 6.9485 | 5.1742 | 0.0928 | 5.267116 | 5.22105 |
| Top SNPs for indirect effects | 4 | Gga_rs316082590 | LOC422264 | 0.7137 | 3.6868 | 0.4754 | 0.5696 | -1.2913 | 0.4505 | -0.84073 | -1.07339 |
| | 4 | Gga_rs313358833 | LOC422265 | 0.6449 | 3.2345 | 0.4310 | 0.5202 | -1.2067 | 0.4235 | -0.78322 | -1.01618 |
| | 4 | Gga_rs314615897 | MAEA | 0.1170 | 2.9505 | 0.0474 | 0.0387 | -0.2799 | 0.3853 | 0.105456 | -0.09807 |
| | 1 | Gga_rs15301842 | – | 0.0393 | 2.9408 | 0.1436 | 0.0149 | -0.1301 | 0.5053 | 0.375199 | 0.050463 |
| | 1 | Gga_rs314551852 | – | 0.0632 | 2.8858 | 0.1100 | 0.0065 | -0.2038 | 0.4994 | 0.295514 | -0.02218 |
| | 1 | Gga_rs317379325 | – | 0.1599 | 2.8473 | 0.0070 | 0.0931 | -0.4789 | 0.5000 | 0.021148 | -0.29321 |
| Overall effects | 14 | Gga_rs313620413 | GRIN2A | 7.4242 | 0.1499 | 9.6599 | 7.4525 | -5.7827 | -0.0498 | -5.83262 | -5.7851 |
| | 1 | Gga_rs314001234 | – | 7.0147 | 1.1067 | 9.0710 | 7.1653 | -26.653 | -0.9018 | -27.5556 | -26.9360 |
| | 7 | Gga_rs315626061 | – | 7.0868 | 0.2220 | 9.0119 | 6.9783 | -22.5681 | 0.2983 | -22.2698 | -22.3520 |
| | 7 | Gga_rs315626061 | – | 6.8300 | 0.3360 | 8.9974 | 6.9529 | 5.1767 | 0.0910 | 5.26783 | 5.2229 |
| | 7 | Gga_rs316509306 | – | 6.8241 | 0.3442 | 8.9952 | 6.9485 | 5.1742 | 0.0928 | 5.267116 | 5.2210 |
| | 7 | Gga_rs15850017 | ZNF385B | 6.6582 | 0.0499 | 8.6397 | 6.6176 | -20.8591 | -0.0718 | -20.9310 | -20.7681 |
| | 7 | Gga_rs316509306 | – | 6.8241 | 0.3442 | 8.9952 | 6.9485 | 5.1742 | 0.0928 | 5.2671 | 5.2210 |
| MTM | 14 | Gga_rs313620413 | GRIN2A | 7.4242 | 0.1499 | 9.6599 | 7.4525 | -5.7827 | -0.0498 | -5.8326 | -5.7851 |
| | 1 | Gga_rs314001234 | – | 7.0147 | 1.1067 | 9.0710 | 7.1653 | -26.6538 | -0.9018 | -27.5556 | -26.936 |
| | 3 | Gga_rs15390496 | EPHA7 | 7.0209 | 0.2214 | 8.6122 | 7.0297 | -22.4233 | -0.2149 | -22.6382 | -22.4098 |
| | 7 | Gga_rs16591372 | OLA1 | 7.0868 | 0.2220 | 9.0119 | 6.9783 | -22.5681 | 0.2983 | -22.2698 | -22.352 |
| | 7 | Gga_rs315626061 | – | 6.8300 | 0.3360 | 8.9974 | 6.9529 | 5.1767 | 0.0910 | 5.26780 | 5.2229 |
| | 7 | Gga_rs316509306 | – | 6.8241 | 0.3442 | 8.9952 | 6.9485 | 5.1742 | 0.0928 | 5.2671 | 5.2210 |
| | 7 | Gga_rs316509306 | – | 6.8241 | 0.3442 | 8.9952 | 6.9485 | 5.1742 | 0.0928 | 5.2671 | 5.2210 |

$ds_{j \rightarrow y(HHP)}$, $inds_{j \rightarrow y(HHP)}$, $TS_{j \rightarrow y(HHP)}$ and $MTMS_{j \rightarrow y(HHP)}$, represents, direct, indirect and overall from SEM and MTM effects of j -th SNP on HHP. The bold values are $-\log_{10}$ (corrected p -value) for each type of significant SNP effects categories.

positive direct effect on BW but may exhibit a negative indirect effect coming from BM, which in turn reduces the total QTL effect on BW. Arguably, the methodology employed here would be most effective when the direct and indirect effects of a QTL are in opposite directions. If the direct and indirect QTL effects are in the same direction, the power of SEM-GWAS may be the same as the overall power of MTM-GWAS. The overall effect ($TS_{j \rightarrow y(HHP)}$) of a given SNP consisted of large indirect ($inds_{j \rightarrow y(HHP)}$) and small direct ($ds_{j \rightarrow y(HHP)}$) effects on HHP, as observed for the top most significant indirect SNPs localized on GGA4 and GAA1, whereas the opposite pattern was observed for the most significant direct SNPs on GAA3, GGA7, and GGA14, which showed large direct and small indirect effects. Although the overall effects of these SNPs from SEM-GWAS and MTM-GWAS were similar, the use of decomposition allowed us to determine that the trait of interest is affected in different manners. For instance, a given SNP effect may largely act directly on HHP without any mediation by BM and BW, whereas another SNP may be transmitting a large effect through a causal path mediated by BM and BW. Collectively, new insight regarding the direction of SNP effects can be obtained using the SEM-GWAS methodology.

DISCUSSION

It is becoming increasingly common to analyze a set of traits simultaneously in GWAS by leveraging genetic correlations between traits (Gao et al., 2014; Wu and Pankow, 2017). In the present study, we illustrated the potential utility of a SEM-based GWAS approach for causal inference and mediation analysis of SNP effects, which has the potential advantage of embedding a pre-inferred causal structure across phenotypic traits (Valente et al., 2010). SEM-GWAS, as an extension of standard MTM, accounts for recursive linking of mediating variables that could be either dependent or independent with restriction on a residual covariance. This is a useful approach when multiple mediators influence the final outcomes via either common or distinct biological pathways (Barfield et al., 2017; Bellavia and Valeri, 2017). SEM-GWAS is achieved by first inferring the structure of networks between phenotypic traits. For this purpose, we used a modified version of the IC algorithm described by Pearl (2009) and modified for implementing in quantitative genetics by Valente et al. (2010). The IC algorithm was used to explore putative causal links among phenotypes obtained from a residual covariance matrix, in a model that accounted for systematic

and genetic confounding factors such as polygenic additive effects. It then produced a posterior distribution of partial residual correlations between any possible pairs of variables. Three different causal path diagrams were inferred from HPD intervals of 0.75, 0.85, and 0.95. We observed that the number of identified paths decreased with an increase in the HPD interval value. Only a path connecting BM and BW was present in all HPD intervals considered. Moreover, we found that the partial residual correlation between BM and HHP was weaker than that between BM and BW. This may explain why the path between BM and HHP was not detected with HPD intervals larger than 0.75.

The primary purpose of estimating the goodness of fit criteria was to determine whether full recursive SEM and MTM models with different assumptions yield the same or nearly the same BIC and AIC scores. Because our results showed that SEM and MTM produced nearly the same goodness of fit criteria, we conclude that the essential difference between these models cannot be articulated in terms of an expressive power of joint distributions or goodness of fit (Valente et al., 2013).

Estimated path coefficients reflect the strength of each causal link, quantifying the proportion of direct and indirect effects of a given SNP or genes on the outcome of interest via the mediator phenotypic traits or the predefined causal pathway between a set of mediators and the target outcome. For instance, a positive path coefficient from BM to BW suggests that a unit increase in BM directly results in an increase in BW. Our results showed that MTM-GWAS and SEM-GWAS were similar in terms of the goodness of fit as per the AIC and BIC criteria. This finding is in agreement with theoretical work of Gianola and Sorensen (2004) and Varona et al. (2007) showing the equivalence between models. Thus, MTM-GWAS and SEM-GWAS produced the same marginal phenotypic distributions and goodness of fit values. A similar approach has been proposed by Li et al. (2006), Mi et al. (2010b), and Wang and van Eeuwijk (2014). The main difference between our approach and theirs is that they used SEM in the context of standard QTL mapping, whereas our SEM-GWAS is developed for GWAS based on a linear mixed model.

The results obtained in this study using the three economic traits in chickens suggest that causal inference and the SEM framework can be used for a set of phenotypes by considering both the raw and partial correlation relationships among traits in breeding programs. For example, in model SEM-A85, BM and HHP are unconditionally independent. However, conditioning on BW results in a non-zero partial correlation. Conditioning on BW breaks the causal chain from BM to HHP as observed in the case of full recursive models (SEM-A75 and SEM-G75) and their partial correlation becomes non-zero. This indicates that when all three variables are causally connected, both raw and partial correlations will all be non-zero, but they will change the magnitude depending on the signs of the path coefficients.

The advantage of SEM-GWAS over MTM-GWAS is that the former decomposes SNP effects by tracing inferred causal networks. Our results showed that by partitioning SNP effects

into direct, indirect, and total components, an alternative perspective of SNP effects can be obtained. As shown in **Table 3** and **Figure 4**, direct and indirect effects may differ in magnitude and sign, acting in the same direction or in an antagonistic manner. Note that the total SNP effects inferred from SEM-GWAS were the same as the estimated SNP effects from MTM-GWAS (**Table 3**). However, knowledge derived from the decomposition of SNP effects may be critical for animal and plant breeders in breaking unfavorable indirect QTL effects by reducing the frequency of undesired alleles or obtaining better SNP effect estimates than those from MTM-GWAS (e.g., Mi et al., 2010b).

CONCLUSION

SEM offers insights into how phenotypic traits relate to each other. We illustrated potential advantages of SEM-GWAS relative to the commonly used standard MTM-GWAS by using three chicken traits as an example. SNP effects pertaining to SEM-GWAS have a different meaning than those in MTM-GWAS. Our results showed that SEM-GWAS enabled the identification of whether a SNP effect is acting directly or indirectly, i.e., mediated, on given trait. In contrast, MTM-GWAS only captures overall genetic effects on traits, which is equivalent to combining direct and indirect SNP effects from SEM-GWAS together. Thus, SEM-GWAS offers more information and provides an alternative view of putative causal networks, enabling a better understanding of the genetic quiddity of traits at the genomic level.

AUTHOR CONTRIBUTIONS

MM carried out the study and wrote the first draft of the manuscript. GR and DG designed the experiment, supervised the study, and critically contributed to the final version of manuscript. GM contributed to the interpretation of results, provided critical insights, and revised the manuscript. BV and AA participated in discussion and reviewed the manuscript. MA, AK, and RM contributed materials and revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

MM wishes to acknowledge the Ministry of Science, Research and Technology of Iran for financially supporting his visit to the University of Wisconsin-Madison. Work was partially supported by the Wisconsin Agriculture Experiment Station under hatch grant 142-PRJ63CV to DG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00455/full#supplementary-material>

REFERENCES

- Anderson, J. C., and Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychol. Bull.* 103:411. doi: 10.1037/0033-2909.103.3.411
- Barfield, R., Shen, J., Just, A. C., Vokonas, P. S., Schwartz, J., Baccarelli, A. A., et al. (2017). Testing for the indirect effect under the null for genome-wide mediation analyses. *Genet. Epidemiol.* 41, 824–833. doi: 10.1002/gepi.22084
- Bellavia, A., and Valeri, L. (2017). Decomposition of the total effect in the presence of multiple mediators and interactions. *Am. J. Epidemiol.* 187, 1311–1318. doi: 10.1093/aje/kwx355
- Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12:232. doi: 10.1186/gb-2011-12-10-232
- Browning, S. R., and Browning, B. L. (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, 1084–1097. doi: 10.1086/521987
- Falconer, D. S., and Mackay, T. F. (1996). Introduction to quantitative genetics (4th edn). *Trends Genet.* 12:280.
- Forni, S., Aguilar, I., and Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* 43:1. doi: 10.1186/1297-9686-43-1
- Gao, H., Zhang, T., Wu, Y., Wu, Y., Jiang, L., Zhan, J., et al. (2014). Multiple-trait genome-wide association study based on principal component analysis for residual covariance matrix. *Heredity* 113, 526–532. doi: 10.1038/hdy.2014.57
- Gianola, D., Fariello, M. I., Naya, H., and Schön, C.-C. (2016). Genome-wide association studies with a genomic relationship matrix: a case study with wheat and arabidopsis. *G3: Genes Genomes Genet.* 6, 3241–3256. doi: 10.1534/g3.116.034256
- Gianola, D., and Sorensen, D. (2004). Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics* 167, 1407–1424. doi: 10.1534/genetics.103.025734
- Hayes, B., and Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome* 53, 876–883. doi: 10.1139/G10-076
- Henderson, C., and Quaas, R. (1976). Multiple trait evaluation using relatives' records. *J. Anim. Sci.* 43, 1188–1197. doi: 10.2527/jas1976.4361188x
- Jamrozik, J., and Schaeffer, L. (2011). Alternative parameterizations of the multiple-trait random regression model for milk yield and somatic cell score via recursive links between phenotypes. *J. Animal Breed. Genet.* 128, 258–266. doi: 10.1111/j.1439-0388.2011.00918.x
- Jiang, G., Chakraborty, A., Wang, Z., Boustani, M., Liu, Y., Skaar, T., et al. (2013). New aQTL SNPs for the CYP2D6 identified by a novel mediation analysis of genome-wide SNP arrays, gene expression arrays, and CYP2D6 activity. *BioMed. Res. Int.* 2013:493019. doi: 10.1155/2013/493019
- Kennedy, B. W., Quinton, M., and van Arendonk, J. A. M. (1992). Estimation of effects of single genes on quantitative traits. *J. Animal Sci.* 70, 2000–2012. doi: 10.2527/1992.7072000x
- Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* 44, 1066–1071. doi: 10.1038/ng.2376
- Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59. doi: 10.1186/1471-2164-14-59
- Kumar, S., and Agrawal, S. (2013). "Disease-oriented causal networks," in *Encyclopedia of Systems Biology* (New York, NY: Springer), 593–594.
- Li, R., Tsaih, S.-W., Shockley, K., Stylianou, I. M., Wergedal, J., Paigen, B., et al. (2006). Structural model analysis of multiple quantitative traits. *PLoS Genet.* 2:e114. doi: 10.1371/journal.pgen.0020114
- Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., and Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nat. Methods* 9, 525–526. doi: 10.1038/nmeth.2037
- Meyer, K. (2007). WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8, 815–821. doi: 10.1631/jzus.2007.B0815
- Meyer, K., and Tier, B. (2012). "SNP Snappy": a strategy for fast genome-wide association studies fitting a full mixed model. *Genetics* 190, 275–277. doi: 10.1534/genetics.111.134841
- Mi, X., Eskridge, K., Wang, D., Baenziger, P. S., Campbell, B. T., Gill, K. S., et al. (2010a). Bayesian mixture structural equation modelling in multiple-trait QTL mapping. *Genet. Res.* 92, 239–250. doi: 10.1017/S0016672310000236
- Mi, X., Eskridge, K., Wang, D., Baenziger, P. S., Campbell, B. T., Gill, K. S., et al. (2010b). Regression-based multi-trait QTL mapping using a structural equation model. *Stat. Appl. Genet. Mol. Biol.* 9:38. doi: 10.2202/1544-6115.1552
- Momen, M., Mehrgardi, A. A., Sheikhy, A., Esmailzadeh, A., Fozi, M. A., Kranis, A., et al. (2017). A predictive assessment of genetic correlations between traits in chickens using markers. *Genet. Sel. Evol.* 49:16. doi: 10.1186/s12711-017-0290-9
- Nock, N., and Zhang, L. (2011). Evaluating aggregate effects of rare and common variants in the 1000 genomes project exon sequencing data using latent variable structural equation modeling. *BMC Proc.* 5 (Suppl. 9):S47. doi: 10.1186/1753-6561-5-S9-S47
- O'Reilly, P. F., Hoggart, C. J., Pomyen, Y., Calboli, F. C., Elliott, P., Jarvelin, M.-R., et al. (2012). MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE* 7:e34861. doi: 10.1371/journal.pone.0034861
- Pearl, J. (2009). Causal inference in statistics: an overview. *Stat. Surveys* 3, 96–146. doi: 10.1214/09-SS057
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* 81, 559–575. doi: 10.1086/519795
- Raykov, T., and Marcoulides, G. A. (2012). *A first Course in Structural Equation Modeling*. Mahwah, NJ: Routledge.
- Rosa, G. J., Valente, B. D., de los Campos, G., Wu, X.-L., Gianola, D., and Silva, M. A. (2011). Inferring causal phenotype networks using structural equation models. *Genet. Sel. Evol.* 43:6. doi: 10.1186/1297-9686-43-6
- Schadt, E. E. (2016). "Chapter 10 - Reconstructing causal network models of human disease A2 - Lehner, Thomas," in *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*, eds B. L. Miller and M. W. State (San Diego, CA: Academic Press), 141–160.
- Sikorska, K., Lesaffre, E., Groenen, P. F., and Eilers, P. H. (2013). GWAS on your notebook: fast semi-parallel linear and logistic regression for genome-wide association studies. *BMC Bioinform.* 14:166. doi: 10.1186/1471-2105-14-166
- Sun, Y., Zhao, G., Liu, R., Zheng, M., Hu, Y., Wu, D., et al. (2013). The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. *BMC Genomics* 14:458. doi: 10.1186/1471-2164-14-458
- Valente, B. D., Rosa, G. J., de los Campos, G., Gianola, D., and Silva, M. A. (2010). Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics* 185, 633–644. doi: 10.1534/genetics.109.112979
- Valente, B. D., Rosa, G. J., Gianola, D., Wu, X.-L., and Weigel, K. (2013). Is structural equation modeling advantageous for the genetic improvement of multiple traits? *Genetics* 194, 561–572. doi: 10.1534/genetics.113.151209
- Van Goor, A., Bolek, K. J., Ashwell, C. M., Persia, M. E., Rothschild, M. F., Schmidt, C. J., et al. (2015). Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. *Genet. Sel. Evol.* 47:96. doi: 10.1186/s12711-015-0176-7
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Varona, L., Sorensen, D., and Thompson, R. (2007). Analysis of litter size and average litter weight in pigs using a recursive model. *Genetics* 177, 1791–1799. doi: 10.1534/genetics.107.077818
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* 94, 73–83. doi: 10.1017/S0016672312000274
- Wang, H., and van Eeuwijk, F. A. (2014). A new method to infer causal phenotype networks using QTL and phenotypic information. *PLoS ONE* 9:e103997. doi: 10.1371/journal.pone.0103997
- Wright, S. (1921). Correlation and causation. *J. Agric. Res.* 20, 557–585.
- Wu, B., and Pankow, J. S. (2017). Genome-wide association test of multiple continuous traits using imputed SNPs. *GBA* 10, 379–386. doi: 10.4310/SII.2017.v10.n3.a2

- Wu, X. L., Heringstad, B., and Gianola, D. (2010). Bayesian structural equation models for inferring relationships between phenotypes: a review of methodology, identifiability, and applications. *J. Anim. Breed. Genet.* 127, 3–15. doi: 10.1111/j.1439-0388.2009.00835.x
- Yang, J., Benyamin, B., McEvoy, B., Gordon, S., Henders, A., and Nyholt, D. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569. doi: 10.1038/ng.608
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Momen, Ayatollahi Mehrgardi, Amiri Roudbar, Kranis, Mercuri Pinto, Valente, Morota, Rosa and Gianola. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Parameters for Resistance to Non-specific Diseases and Production Traits Measured in Challenging and Selection Environments; Application to a Rabbit Case

Mélanie Gunia^{1*}, Ingrid David¹, Jacques Hurtaud², Mickaël Maupin², Héléne Gilbert¹ and Hervé Garreau¹

¹ GenPhySE, INRA, ENVT, Université de Toulouse, Castanet Tolosan, France, ² HYPHARM SAS, La Corbière, Roussay, Sèvres, France

OPEN ACCESS

Edited by:

Andrea B. Doeschl-Wilson,
Roslin Institute, University of
Edinburgh, United Kingdom

Reviewed by:

Allan Schinckel,
Purdue University, United States
Adriana Santana Carmo,
Deoxi Biotecnologia, Brazil

*Correspondence:

Mélanie Gunia
melanie.gunia@inra.fr

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 23 April 2018

Accepted: 24 September 2018

Published: 16 October 2018

Citation:

Gunia M, David I, Hurtaud J,
Maupin M, Gilbert H and Garreau H
(2018) Genetic Parameters for
Resistance to Non-specific Diseases
and Production Traits Measured in
Challenging and Selection
Environments; Application to a Rabbit
Case. *Front. Genet.* 9:467.
doi: 10.3389/fgene.2018.00467

Breeding for disease resistance is a challenging but increasingly necessary objective to overcome the issues with the reduced use of antibiotics and growing concern for animal welfare while limiting economic losses. However, implementing such strategies is a complex process because animals face numerous diseases, and the environments on selection farms differ from those on commercial farms. We evaluated whether selection for resistance to non-specific diseases based on a single visual record in selection (S) and challenging (Ch) environments is possible. Records from 23,773 purebred rabbits born between 2012 and 2016 were used in this study. After weaning (at 32 days of age), 17,712 rabbits were raised in the S environment and 6,061 sibs were raised in the Ch environment. Clinical signs of disease were recorded for all animals at the end of the test, at a single time point, at 70 or 80 days of age. The causes of mortality occurring before the end of the test were also recorded. Three disease traits were analyzed: signs of respiratory disease, signs of digestive disease, and a composite trait (Resist) taking into account signs of digestive, respiratory and various infectious diseases. This latter composite trait is proposed to capture the global resistance to disease. All disease traits were binary, with 0 being the absence of symptoms. Two production traits were also recorded: the number of kits born alive (4,121 litters) and the weaning weight (13,090 rabbits). Disease traits were analyzed with animal threshold models, assuming that traits are different in the two environments. Bivariate analyses were carried out using linear animal models. The heritabilities of the disease traits ranged from 0.04 ± 0.01 to 0.11 ± 0.03 . The genetic correlations between disease traits in both environments were below unity (≤ 0.84), indicating genotype by environment interactions. Most of the genetic correlations between disease and production traits were not significantly different from

zero, except between the weaning weight and Resist_S, with a favorable correlation of -0.34 ± 0.12 . Given these genetic parameters, for the same level of exposure of rabbits to pathogens, the expected response to selection is a reduction of disease incidence of 4–6% per generation.

Keywords: heritability, genetics, resistance to disease, farming, rabbit, genetic parameters, genotype-environment interaction

INTRODUCTION

Breeding for disease resistance is becoming increasingly important to reduce the use of antibiotics and address the growing concern for animal welfare. It also contributes to reducing production costs at both the selection and commercial levels (Phocas et al., 2016). Improving disease resistance by selection is challenging. During their lifetime, animals can face various pathogens, many of which are not always identified. In addition, when the selection environment differs considerably from commercial environments (i.e., the higher biosecurity level of selection environments entails a lower expression of disease) little or no selection pressure is applied on this trait. To implement such selection, there is still a need for phenotypes that can be easily measured, at a reasonable cost (Merks et al., 2012). In rabbits, previous studies have shown that simple health records, measured once on growing animals of the selection nucleus, can be used to improve disease resistance (Eady et al., 2007; Garreau et al., 2012; Gunia et al., 2015). However, it is not known if such selection will be beneficial for maintaining the health and productivity of animals reared in commercial conditions. To address this issue, disease symptoms were recorded in rabbits at Hypharm's facilities in both a selection environment and more challenging environments. The aim of this study was to determine whether selection for resistance to non-specific diseases is possible, and how records from different environments can improve the genetic gains. We first estimated the genetic parameters of disease resistance traits in two contrasting environments and their genetic and phenotypic correlations with the main production traits. Then, we assessed the expected genetic progress for disease resistance for various selection strategies, including records from different environments in the genetic evaluations.

MATERIALS AND METHODS

This study was carried out in accordance with the national regulations of agriculture in the framework of the selective breeding schemes of the Hypharm breeding company.

Animals

Data were collected for animals of the AGP77 maternal rabbit line (Hypharm, Roussay, France). Records from 23,773 purebred rabbits born between 2012 and 2016 were analyzed. This line was created from rabbits of the New Zealand breed in 1975. Animals have been selected for litter size and weaning weight (direct and maternal effects) since 2002. In our dataset, does were inseminated every 42 days and kits were weaned at 32

days of age. All rabbits were born and weaned on a nucleus farm, i.e., in a highly bio-secure and controlled environment, where all candidates for selection were further tested. This farm is hereafter referred to as the selection environment (S). After being weighed at weaning, these rabbits were reared in two different environments: (1) the S environment or (2) sib-testing farms, which were 3 farms with less favorable sanitary conditions than the nucleus farm. The sib-testing farms are hereafter referred to as the challenging environment (Ch). Some full sibs and half sibs of the candidates for selection were tested in Ch. The aim of the challenging environment was to mimic the less protected environmental conditions encountered on some commercial rabbit farms. Farms were semi-open rabbit farms with no artificial heating, cooling or ventilation systems. Rabbits of various age classes were reared in the same rooms, and the frequency of veterinary treatments was kept to a minimum. Sick rabbit groups were treated with water medication according to veterinary requirements. Rabbits (mostly males) from every second weaning batch were sent to the Ch environment at weaning. In total, 6,061 rabbits had health records in the Ch environment (5,864 males and 197 females) and 17,712 in the S environment (5,499 males and 12,213 females). The pedigree included 332 sires and 849 dams. A total of 228 sires had kits with health records in both environment, 29 sires in S only and 1 sire in Ch only.

Traits

Clinical signs of diseases occurring naturally on farms were recorded at a single time point, at the end of the test at 70 or 80 days of age. Very mild clinical signs of diseases were recorded. The most likely cause of death was also recorded after necropsy for rabbits that died between weaning and the end of the test. Disease traits were coded as 0 (absence) or 1 (disorder = morbidity at the end of the test or mortality between weaning and the end of the test). Clinical sign of diseases were not recorded between weaning and the end of the test. Rabbits categorized at healthy (0) at 70 or 80 days of age might have been sick individuals who had recovered. Disorders were further divided into the following categories: (1) digestive disease (Dig), which included diarrhea, bloated abdomen, and any form of digestive symptoms, (2) respiratory disease (Resp) which included nasal discharge, lung lesions, eye infection, wry neck, and (3) non-specific diseases (Resist), which combined Dig, Respi, abnormally low weight, and other clinical signs of infectious origin. The disease traits were treated as separate traits depending on the environment (S or Ch), resulting in a total of 6 disease traits: Dig_S, Resp_S, Resist_S, Dig_Ch, Resp_Ch,

TABLE 1 | Total prevalence (in %) of non-specific diseases (Resist), respiratory (Resp) and digestive (Dig) diseases in selection (S) and challenging (Ch) environments from 2012 to 2016¹.

| | S | Ch |
|--------|----|----|
| Resist | 26 | 41 |
| Resp | 14 | 24 |
| Dig | 11 | 16 |

¹ *N* = 17,712 in S and *N* = 6,061 in Ch.² Very mild clinical signs of a disease were taken into account.**TABLE 2** | Number of records (*N*), mean and standard deviation (Std) for the number of kits born alive (NBA) and weaning weight (WW).

| | N | Mean | Std |
|--------|---------------------|------|------|
| NBA | 4,121 ¹ | 9.92 | 3.34 |
| WW (g) | 13,090 ² | 664 | 102 |

¹ Number of litters, ² Number of rabbits.

Resist_Ch. The production traits were the number of kits born alive per doe (NBA) and the weaning weight (WW), which were exclusively recorded in the S environment. Descriptive statistics of the disease and production traits are listed in **Tables 1, 2**.

Genetic Parameters Analyses

All traits were analyzed using a restricted maximum likelihood method, with the ASReml 3.0 software (Gilmour et al., 2009). Variance components and heritabilities were estimated using single-trait animal threshold models with a logit link function for the binary disease traits and multiple trait linear animal models for the production traits. Genetic and phenotypic correlations between traits were estimated using multiple trait linear animal models. The models included a random additive polygenic effect for all traits, a random common litter effect for the disease resistance traits and WW, a random maternal environmental effect and a random maternal genetic effect for WW, and a permanent environmental effect to account for the repeated measurement of NBA for does. For the 4% of kits cross-fostered at birth to another doe, the maternal genetic effect, the common litter effect, and the maternal environmental effects were assigned to the adoptive suckling mother. The significance of the fixed effects was determined for each trait using the Wald F statistic, which is similar to an ANOVA (Gilmour et al., 2009). Fixed effects were first tested together, and then a stepwise selection of the significant effects was applied. Significant fixed effects ($P < 0.05$) were maintained in subsequent analyses (**Table 3**). They were: batch and sex for the disease traits and WW, parity of the dam for WW, and a Ch farms effect for the disease traits measured in Ch. The combined effects of year-season of kitting and parity-physiological status (lactating or not at insemination) were applied for NBA.

To estimate genetic correlations, we used analysis methods for continuous data, which are not theoretically optimal. The suitable methodology is the threshold model (Gianola, 1982). However, assumption of a continuous distribution for these

TABLE 3 | Fixed effects included in the models (x), not significant (NS), or not tested (-) for non-specific diseases (Resist), respiratory (Resp), and digestive (Dig) disease in selection (S) and challenging (Ch) environments, for weaning weight (WW), and number of kits born alive (NBA).

| Effect | N of levels | Resist_S Resp_S Dig_S | Resist_Ch Resp_Ch Dig_Ch | WW | NBA |
|--|-------------|-----------------------------|--------------------------------|----|-----|
| Batch | 29 | x | x | x | - |
| Sex | 2 | x | x | x | - |
| Ch farms | 3 | - | x | - | - |
| Parity of the dam | 5 | NS | NS | x | - |
| Weaning age | 3 | NS | NS | NS | - |
| Year*season of kitting | 15 | - | - | - | x |
| Parity*physiological status ¹ | 9 | - | - | - | x |

¹ Lactating or not at insemination.

traits is justified for genetic evaluation and for estimates of genetic correlations with continuous traits (Kadarmideen et al., 2003). Several studies showed that the estimates of heritability or breeding values from linear and threshold models are highly correlated (Matos et al., 1997a,b; Ramirez-Valverde et al., 2001). The difference between these methodologies has been shown to be negligible when the incidence of the binary response was between 25 and 75% (Meijering and Gianola, 1985). Except in the case where fixed effects were added, nested models were compared using the restricted likelihood ratio test. When the model comparison corresponded to a test of parameter on the boundary of parameter space (test of variance different from 0 or test of correlation different from 1), the distribution of this test statistic under the null hypothesis is a 50:50 mixture of χ^2_q and χ^2_{q+1} distributions (Morrell, 1998), where, *q* is the number of random effects in the reduced model (residual effect excluded). To obtain the standard error of the heritability, we performed a multivariate sampling approach of the variance components as described by Houle and Meyer (2015) 10,000 times and computed heritability for each sample. Standard deviation of the sampling was then used as an estimation of the standard error of the heritability. In addition, we checked the normality of the distribution obtained. If the hypothesis of normality was not rejected, we then used a Student's *t*-test to compare heritabilities obtained in the S and Ch environments.

Simulation of Breeding Schemes

To illustrate the genetic progress that could be obtained for non-specific disease resistance, we tested various breeding schemes including Resist_S or Resist_Ch.

General Parameters

We used the deterministic simulation program SelAction 2.2 (Rutten et al., 2002) to compare the expected selection responses for breeding schemes including resistance to non-specific disease. SelAction predicts responses to selection on pseudo-BLUP estimated breeding values. In our study, we used the option of discrete generations and 1-stage selection. We simulated a selection nucleus of 140 dams and 35 sires. We assumed 7

progeny per litter with a sex ratio of ½. The selection intensity was 15% for males and 25% for females. Rabbits were selected at 70 days of age. The genetic parameters used for the simulation were those obtained in the first part of the study and the variance components obtained with a linear model for Resist_S and Resist_Ch.

Alternative Breeding Schemes

We compared expected selection responses for fictive breeding objectives including disease resistance in different ways. The four tested breeding objectives were:

$$\begin{aligned} \text{HResist_S_Ch} &= 3 \times A_{\text{NBA}} + 0.15 \times A_{\text{WW_direct}} + 0.15 \times A_{\text{WW_maternal}} - 65 \times A_{\text{Resist_S}} - 65 \times A_{\text{Resist_Ch}} \\ \text{HResist_S} &= 3 \times A_{\text{NBA}} + 0.15 \times A_{\text{WW_direct}} + 0.15 \times A_{\text{WW_maternal}} - 130 \times A_{\text{Resist_S}} \\ \text{HResist_Ch} &= 3 \times A_{\text{NBA}} + 0.15 \times A_{\text{WW_direct}} + 0.15 \times A_{\text{WW_maternal}} - 130 \times A_{\text{Resist_Ch}} \\ \text{HProduction} &= 3 \times A_{\text{NBA}} + 0.15 \times A_{\text{WW_direct}} + 0.15 \times A_{\text{WW_maternal}} \end{aligned}$$

with A_x denoting the true breeding value for trait X . The traits NBA was expressed in number of kits, WW_direct and WW_maternal in g, and Resists_S and Resist_Ch in %. The corresponding weights were given in euros per physical unit of the traits.

The weights did not have any economic meaning, as they were derived by trials and errors based on the desired gain methodology (Brascamp, 1984). Pre-trial simulations with SelAction (Rutten et al., 2002) based on a breeding objective including all traits were run to estimate the genetic gain for various sets of weights. Then, a set of weights was arbitrarily chosen to improve non-specific disease resistance in both S and Ch, while increasing the weaning weight and maintaining a stable NBA, resulting in the HResist_S_Ch breeding objective presented above.

For comparison reason, we also assessed the selection response for a breeding objective including only production traits (HProduction) by using the same weights on the production traits. We studied the indirect expected selection response for

each trait, even though it was excluded from the breeding objective, as for instance response in Resist_S and Resist_Ch, when selection occurred for HProduction.

For each breeding objective including Resist_S or Resist_Ch, Resist was recorded with one or the other of the following modalities:

- 1) With Resist_Ch records: Resist_Ch was recorded on a quarter of the sibs of the selection candidate, while Resist_S was recorded on the selection candidates and on the remaining three-quarters of the sibs.
- 2) Without Resist_Ch records: all selection candidates and sibs had records for Resist_S only. Resist_Ch was not recorded.

For HProduction, we considered that disease resistance traits were not recorded.

Correlations Between Breeding Objectives

The correlations between breeding objectives were calculated as:

$$\begin{aligned} \text{Covariance}(H_i, H_j) &= \mathbf{W}_i^T \times \text{Var}(A_{ij}) \times \mathbf{W}_j \\ \text{Correlation}(H_i, H_j) &= \frac{\text{Covariance}(H_i, H_j)}{\sqrt{\sigma_{H_i}^2 \times \sigma_{H_j}^2}} \end{aligned}$$

with \mathbf{W}_i^T being the row vector ($1 \times n$) of the weights of the breeding objective H_i , \mathbf{W}_j the column vector ($m \times 1$) of the weights of the breeding objective H_j , $\text{Var}(A_{ij})$ the genetic variance-covariance matrix ($n \times m$) of the traits in the breeding objective H_i and H_j , and $\sigma_{H_i}^2$ the variance of the breeding objective.

RESULTS

Phenotypes

The phenotypes analyzed are shown in **Tables 1, 2**. The moderately high disease prevalence over the test period (26% for Resist_S and 41% for Resist_Ch) reflects the accurate recording of even the slightest clinical signs of disease. Disease prevalence was higher in Ch than S: 5 percentage points for Dig, 10 percentage

TABLE 4 | Estimates of variance components, heritabilities, common litter effect, and genetic correlations between selection (S) and challenging (Ch) environments for non-specific disease (Resist), respiratory (Resp), and digestive (Dig) disease (\pm standard errors).

| Environment | Resist | | Resp | | Dig | |
|--------------------------------|-----------------------------------|-----------------|-----------------|-----------------|-----------------------------------|-----------------|
| | S | Ch | S | Ch | S | Ch |
| σ_a^2 | 0.15 \pm 0.04 | 0.29 \pm 0.08 | 0.26 \pm 0.06 | 0.31 \pm 0.08 | 0.26 \pm 0.07 | 0.42 \pm 0.12 |
| $\sigma_{\text{com.litter}}^2$ | 0.14 \pm 0.03 | 0.01 \pm 0.01 | 0.19 \pm 0.04 | 0.00 \pm 0.00 | 0.20 \pm 0.05 | 0.04 \pm 0.07 |
| σ_e^2 | 1.00 \pm 0.00 | 1.00 \pm 0.00 | 1.00 \pm 0.00 | 1.00 \pm 0.00 | 1.00 \pm 0.00 | 1.00 \pm 0.00 |
| σ_p^2 | 3.59 \pm 0.04 | 3.60 \pm 0.07 | 3.75 \pm 0.06 | 3.60 \pm 0.08 | 3.74 \pm 0.07 | 3.75 \pm 0.10 |
| h^2 | 0.04 \pm 0.01 | 0.08 \pm 0.02 | 0.07 \pm 0.02 | 0.09 \pm 0.02 | 0.07 \pm 0.02 | 0.11 \pm 0.03 |
| $c^2_{\text{com.litter}}$ | 0.04 \pm 0.01 | 0.01 \pm 0.01 | 0.05 \pm 0.01 | 0.00 \pm 0.00 | 0.05 \pm 0.01 | 0.01 \pm 0.02 |
| r_g^1 | 0.70 \pm 0.13 | | 0.84 \pm 0.12 | | 0.48 \pm 0.16 | |

σ_a^2 , direct genetic variance; $\sigma_{\text{com.litter}}^2$, common litter effect variance; σ_e^2 , residual variance; σ_p^2 , phenotypic variance; h^2 , direct heritability; $c^2_{\text{com.litter}}$, common litter effect; r_g , genetic correlation. Variance components; heritability and common litter effect were estimated with a single-trait animal threshold model (logit transformation). Genetic correlations between S and Ch were estimated with a two-trait linear animal model.

¹Genetic correlation values in bold type are significant different from one at $P < 0.001$.

TABLE 5 | Genetic correlations (above diagonal) and phenotypic correlations (below diagonal) for non-specific disease (Resist), respiratory (Resp), and digestive disease (Dig) in selection (S) and challenging (Ch) environments (\pm standard errors).

| | Resist_S | Resp_S | Dig_S | | Resist_Ch | Resp_Ch | Dig_Ch |
|----------|-----------------------------------|------------------------------------|-----------------------------------|-----------|-----------------------------------|------------------------------------|-----------------------------------|
| Resist_S | | 0.69 \pm 0.08 | 0.66 \pm 0.09 | Resist_Ch | | 0.74 \pm 0.08 | 0.52 \pm 0.13 |
| Resp_S | 0.67 \pm 0.00 | | −0.06 \pm 0.16 | Resp_Ch | 0.66 \pm 0.01 | | −0.19 \pm 0.17 |
| Dig_S | 0.60 \pm 0.01 | −0.12 \pm 0.01 | | Dig_Ch | 0.54 \pm 0.01 | −0.25 \pm 0.01 | |

Correlations were estimated with two-trait linear animal models.

Values in bold type are significantly different from zero at $P < 0.01$.

points for Resp, and 15 percentage points for Resist. The average NBA was 9.92 ± 3.34 kits, and WW was 664 ± 102 g.

Genetic Parameters of Disease Resistance Traits in Selection and Challenging Environments

The genetic parameters of the disease traits are provided in **Table 4**. The genetic correlations between environments for each disease trait are given in **Table 5**. The heritabilities of the disease traits were low, ranging from 0.04 ± 0.01 (Resist_S) to 0.11 ± 0.03 (Dig_Ch). Heritabilities tended to be higher in Ch than in S. However, the difference was not significantly different from zero. The common litter effect also tended to be higher in S (0.05 ± 0.01) than in Ch (0.01 ± 0.01). The genetic and phenotypic correlations between Resist on the one hand, and Dig and Resp on the other hand, were moderate to high, ranging from 0.52 ± 0.13 to 0.74 ± 0.08 , and were similar in both environments (**Table 5**). The phenotypic correlation between Resp and Dig was moderate and negative, while the genetic correlation was negative but not significantly different from zero. The genetic correlations between Ch and S for each disease resistance trait were below unity, demonstrating significant interaction between the genotype and the environment for Resist and Dig, but not for Resp (**Table 4**). The genetic correlation between S and Ch environments was higher for Resp (0.84 ± 0.12) than for Dig (0.48 ± 0.16), the estimate for the composite trait Resist being intermediate (0.70 ± 0.13).

Genetic Parameters of Production Traits and Correlations With Disease Traits

The variance component estimates of the production traits are shown in **Table 6** and the correlations with the disease resistance traits in **Table 7**. The heritability of NBA was low (0.16 ± 0.03); the direct heritability of WW was moderate (0.29 ± 0.04) and its maternal heritability was very low (0.05 ± 0.02). The phenotypic correlation between NBA and WW was null. The genetic correlation between NBA and direct effects for WW was negative but not significantly different from zero. On the contrary, the genetic correlation between NBA and WW_maternal was moderate and positive (0.51 ± 0.16). The phenotypic correlations between Resist_S or Resist_Ch and the production traits were negative and low to moderate. Most of the genetic correlations between Resist_S or Resist_Ch and the production traits were not significantly different from 0, except for Resist_S and WW_direct which was negative (i.e., favorable).

TABLE 6 | Estimates of variance components, heritabilities, common litter effect, permanent environment effect for the number of kits born alive (NBA) and weaning weight (WW) (\pm standard errors).

| | NBA | WW |
|-------------------------|------------------|-----------------|
| σ^2_a | 1.71 ± 0.38 | 2986 ± 420 |
| $\sigma^2_{com.litter}$ | | 1218 ± 104 |
| $\sigma^2_{mat.env}$ | | 730 ± 140 |
| $\sigma^2_{perm.env}$ | 0.32 ± 0.23 | |
| σ^2_m | | 474 ± 168 |
| σ^2_e | 8.47 ± 0.21 | 5620 ± 226 |
| σ^2_p | 10.49 ± 0.30 | 10186 ± 202 |
| h^2 | 0.16 ± 0.03 | 0.29 ± 0.04 |
| $c^2_{com.litter}$ | | 0.12 ± 0.01 |
| $c^2_{mat.env}$ | | 0.07 ± 0.01 |
| $c^2_{perm.env}$ | 0.03 ± 0.02 | |
| m^2 | | 0.05 ± 0.02 |
| Repeatability | 0.19 ± 0.02 | |

σ^2_a , direct genetic variance; $\sigma^2_{com.litter}$, common litter effect variance; $\sigma^2_{mat.env}$, maternal environment variance; $\sigma^2_{perm.env}$, permanent environmental variance; σ^2_m , maternal genetic variance; σ^2_e , residual variance; σ^2_p , phenotypic variance; h^2 , direct heritability; $c^2_{com.litter}$, common litter effect; $c^2_{mat.env}$, maternal environment effect; m^2 , maternal heritability; repeatability, $(\sigma^2_a + \sigma^2_{perm.env}) / \sigma^2_p$. All variance component were estimated with three-trait linear animal models.

Expected Genetic Gain

We compared the selection response for four breeding objectives including Resist_S, Resist_Ch, or both, or only production traits. The breeding objectives were highly correlated, with correlations greater than or equal to 0.75 (**Table 8**). The correlation was very high between the breeding objectives with Resist_S or Resist_Ch on the one hand and the breeding objective including both traits on the other hand (0.93).

The expected genetic gain is provided in **Table 9** for the four breeding objectives and the two recording modalities for Resist_Ch. On average across scenarios, the expected genetic gain per generation expressed in genetic standard deviation was -0.45 for Resist_S, -0.27 for Resist_Ch, -0.02 for NBA, 0.44 for WW_dir, and -0.19 for WW_mat. The selection response was very stable for Resist_S across the breeding schemes that included Resist_S or Resist_Ch in the breeding objective, with values ranging from -4.5 to -4.8% . The selection response increased for Resist_Ch along with the weight given to this trait in the breeding objective. The highest selection response for both Resist_S and Resist_Ch was obtained for the breeding

TABLE 7 | Genetic correlations (above diagonal) and phenotypic correlations (below diagonal) for non-specific diseases (Resist) in selection (S) or challenging (Ch) environments, number of kits born alive (NBA), and the direct and maternal effects of weaning weight (WW) (\pm standard errors).

| | Resist_S | Resist_Ch | NBA | WW_direct | WW_maternal |
|-----------|------------------------------------|------------------------------------|------------------|------------------------------------|-----------------------------------|
| Resist_S | | 0.70 \pm 0.13 | -0.08 ± 0.14 | -0.34 ± 0.12 | -0.06 ± 0.20 |
| Resist_Ch | — | | -0.06 ± 0.16 | -0.05 ± 0.14 | -0.04 ± 0.22 |
| NBA | -0.38 ± 0.03 | -0.37 ± 0.02 | | -0.22 ± 0.11 | 0.51 \pm 0.16 |
| WW_direct | -0.13 ± 0.01 | -0.11 ± 0.02 | 0.03 ± 0.03 | | 0.71 \pm 0.12 |

Correlations were estimated with three-trait linear animal models.
Values in bold type are significantly different from zero at $P < 0.05$.

TABLE 8 | Correlations between the breeding objectives¹.

| | HResist_Ch | HResist_S_Ch | HProduction |
|-------------|------------|--------------|-------------|
| HResist_S | 0.75 | 0.93 | 0.76 |
| HResist_Ch | | 0.93 | 0.78 |
| HProduction | | | 0.77 |

¹HProduction, breeding objective including the direct and maternal component of weaning weight and the number of kits born alive.

HResist_S, HProduction + Resistance for non-specific diseases in the selection environment

HResist_Ch, HProduction + Resistance for non-specific diseases in the challenging environment

HResist_S_Ch, HProduction + Resistance for non-specific diseases in the selection and in the challenging environment.

scheme Production + Resist_Ch with records for Resist_Ch, with -4.8% for Resist_S and -5.1% for Resist_Ch. For each breeding objective including a disease resistance trait, we quantified the expected genetic gain obtained using records from the challenging environment. Recording Resist in the challenging environment (on a quarter of the sibs of the selection candidates) always improved the selection response for Resist_Ch for all the breeding objectives. This additional genetic progress led to a reduction of disease incidence for Resist_Ch of 1 percentage point per generation (on average across scenarios), which represents an additional genetic gain of 25% (compared with the scenarios where Resist_Ch was not recorded).

The correlated selection response for disease resistance traits was low but still favorable when these traits were not included in the breeding objective and not recorded (-2.6% for Resist_S and -0.6% for Resist_Ch with HProduction). HProduction had the highest genetic gain for WW_direct (35 g), but the lowest for all the other traits. Unfavorable trends were obtained for WW_mat with all scenarios (-4.3 g on average), but they were lower (-2.2 g) for the scheme HResist_Ch with records for Resist_Ch. However, there was less genetic gain for WW_dir with this scenario (16.5 g). NBA was also more stable for this scenario with a very slight increase of the trait (0.005 kits).

DISCUSSION

Improving resistance to non-specific diseases seems to be possible in both selection and more challenging environments.

Resistance to non-specific diseases is a heritable trait in both environments.

Breeding for Disease Resistance or Tolerance to Non-specific Diseases?

In this study, we considered that an animal was affected with a disease if it showed at least one clinical symptom of infection at a single time point. Such observations of animals under normal or more challenging production conditions are a simple and direct approach in order to select for genetic resistance. However, the expression of resistance to disease is questionable (Rothschild, 1998). Sick animals without symptoms may have been categorized as healthy (poor sensitivity) while healthy rabbits or recovering rabbits may have been categorized as sick animals (poor specificity). This could lead to an underestimation of heritability (Bishop and Woolliams, 2010). The $\ll \text{true} \gg$ heritability of disease resistance is likely to be higher than our estimates. Nevertheless, the measure proposed here for Resist is simple, easy to record, can be routinely collected on farms, and seems to be a good proxy for improving the resistance and tolerance to the most common diseases faced by animals in various production conditions. No experimental challenges are required and no particular disease is given priority over another. As emphasized by various authors (Guy et al., 2012; Merks et al., 2012), the main issue for effective genetic selection for disease resistance is the identification of phenotypes that can be easily measured and routinely collected on farms. In our manuscript, we use the terms “selection for disease resistance” in a very general way to qualify the genetic improvement of rabbit health. However, it may include both resistance and tolerance to disease. Host resistance refers to the ability to reduce pathogen replication within a host, whereas, tolerance refers to the ability to reduce the impact of pathogens on host performance without necessarily affecting the pathogen burden (Doeschl-Wilson and Kyriazakis, 2012). Tolerance can also be more broadly assessed against abiotic factors (temperature) or production diseases (Kause and Ødegård, 2012). Our phenotypes were based on observed clinical signs at a definite time. They included no information about the presence of pathogens and the infection dynamic, if any. Recording the pathogen burden seems unfeasible on farms due to the high number of pathogen types and the cost of such analyses when trying to limit production costs. The healthy phenotype in our study could therefore be an expression of resistance (the animals succeed

TABLE 9 | Expected direct selection responses or correlated responses to selection per generation (10 months) in trait unit for non-specific disease resistance (Resist) in the selection (S) or challenging (Ch) environment, for number of kits born alive (NBA), and for the direct and maternal components of weaning weight (WW) for four alternative breeding objectives¹ including or not records for Resist_Ch.

| Breeding objective | Breeding schemes | | | | | | |
|--------------------|------------------|--------|------------|--------|--------------|--------|-----------------|
| | HResist_S | | HResist_Ch | | HResist_S_Ch | | HProduction |
| | Yes | No | Yes | No | Yes | No | No ² |
| Resist_S (%) | −4.6 | −4.5 | −4.6 | −4.7 | −4.8 | −4.7 | −2.6 |
| Resist_Ch (%) | −4.2 | −3.6 | −5.7 | −4.4 | −5.1 | −4.0 | −0.6 |
| NBA (n of kits) | −0.032 | −0.033 | 0.005 | 0.001 | −0.013 | −0.017 | −0.118 |
| WW_direct (g) | 26.345 | 26.464 | 16.471 | 19.222 | 21.892 | 23.233 | 34.990 |
| WW_maternal (g) | −4.708 | −4.626 | −2.225 | −2.429 | −3.535 | −3.616 | −8.720 |

¹HProduction, breeding objective including the direct and maternal component of weaning weight and the number of kits born alive.

HResist_S, HProduction + Resistance for non-specific diseases in the selection environment.

HResist_Ch, HProduction + Resistance for non-specific diseases in the challenging environment.

HResist_S_Ch, HProduction + Resistance for non-specific diseases in the selection and in the challenging environment.

²NBA, WW_direct and WW_maternal are the only traits recorded for this breeding scheme.

to reduce pathogen replication), tolerance to pathogens (the animal is a carrier of the pathogen but maintains its level of performance without showing clinical signs), or tolerance to metabolic disorders (e.g., noninfectious digestive disorders), or result from the absence of contact with pathogens. It has been argued that disease resistance mechanisms are often pathogen-specific, while tolerance mechanisms that prevent or repair damage may be more host than pathogen specific, and may thus offer generic protection for a range of pathogens (Doeschl-Wilson and Kyriazakis, 2012). In our case, we may be improving tolerance, and we may also be improving resistance by selecting animals with a more efficient innate immune response. As observed by Glass (2012), “distinct host resistance and tolerance traits may be less common than traits that involve elements of both strategies which are likely to have evolved together to overcome infectious threats.”

Heritability of Resistance to Non-specific Diseases

The heritability of the disease traits was low. Similar heritabilities were found in French paternal rabbit lines (Gunia et al., 2015) with heritabilities ranging from 0.030 ± 0.003 to 0.041 ± 0.004 for disease traits on the underlying scale. Other studies estimated higher heritability on the observed scale for disease traits: 0.12 ± 0.05 for bacterial infections in Australian rabbits (Eady et al., 2007), 0.17 ± 0.09 to 0.30 ± 0.06 for non-specific mortality, respiratory diseases, and epizootic rabbit enteropathy in Spanish paternal rabbit lines (Ragab et al., 2015). We found no genetic correlations between respiratory and digestive diseases. This result confirms previous results obtained in paternal lines (Gunia et al., 2015). Nonetheless, contrary to the previous study in which only the main disease syndrome was recorded, two disease syndromes could be recorded for each animal in the present dataset. This reduced the bias caused by the limited recording of multiple disease symptoms for the same animal. However, if the two main syndromes were from the same kind of disease (either digestive or respiratory), a third syndrome from another

kind of disease would not be registered. The absence of genetic correlation could depend partially from this recording method. The composite trait Resist was genetically correlated with Resp and Dig, and the correlations were similar in the Ch and S environments. Resist could therefore be a good indicator for improving general disease resistance and reducing the sensitivity of rabbits to digestive and respiratory diseases.

Genotype by Environment Interactions for Disease Resistance Traits

Genotype by Environment ($G \times E$) interactions were demonstrated for Resist and Dig, with genetic correlations significantly below unity between S and Ch. We also observed a scaling effect, with higher genetic variances in Ch than S. They were compensated by lower variances due to common litter environment, leading to very similar total variance between S and Ch. The lower between-environment genetic correlation observed for Dig (0.48 ± 0.16) compared with Resp (0.84 ± 0.12) may be explained by the higher variability of digestive diseases compared with respiratory diseases. Digestive syndromes can be caused by various pathogens and give rise to various diseases (epizootic rabbit enteropathy, coccidiosis, enterotoxaemia, colibacillosis) that differ among challenging environment farms, whereas respiratory syndromes are mainly caused by *Pasteurella multocida* in rabbits. In pigs, $G \times E$ interactions have been described for production traits between nucleus and testing farms (Merks, 1989), and even between farms of good health status (Hermesch et al., 2015). Another study in rabbits reported $G \times E$ interactions for nonspecific mortality, respiratory diseases, and epizootic rabbit enteropathy between animals fed *ad libitum* and a restricted diet (Ragab et al., 2015) with genetic correlations for the disease traits between the two feeding systems ranging from 0.26 ± 0.09 to 0.68 ± 0.07 . In poultry, genetic correlations ranging from 0.78 to 0.82 were observed for footpad dermatitis of broilers reared in two contrasting environments. $G \times E$ interactions for disease resistance have also been reported for different times of infection occurring naturally

in farms, for example resistance and resilience from low to high worm challenges in sheep (Riley and Van Wyk, 2009). G×E interactions have been observed for reproduction traits during high and low challenge loads (due to natural disease agent and other stressors) in pig farms (Herrero-Medrano et al., 2015), or before and after an outbreak of Porcine Reproductive and Respiratory Syndrome (Lewis et al., 2008).

Various ways to account for G×E interactions in selection have been proposed depending on the magnitude of the genetic correlation of traits evaluated between environments. Robertson (1959) proposed that genetic correlations below the threshold of 0.80 could be considered as having biological importance, with significant reranking of animals occurring across environments. Alternative breeding strategies can therefore be considered. Simulations in dairy cattle have shown that developing specific breeding programs for each environment is interesting when the genetic correlation falls below 0.61 (Mulder et al., 2006). The genetic correlation for Resist_Ch and Resist_S is 0.70 ± 0.13 , which means that the application of a common breeding program in both environments to improve Resist could be an appropriate strategy. However, rabbit breeding schemes differ from those for dairy cattle: they are pyramidal (as in poultry and pigs) without progeny testing. Therefore, further research needs to be undertaken to compare the advantages of running common or separate breeding schemes for pyramidal selection.

Disease Resistance and Production Traits

Estimates of variance components for production traits were generally consistent with the literature (Garcia and Baselga, 2002; Mocé and Santacreu, 2010; Loussouarn et al., 2012; David et al., 2015). The correlations between Resist and the production traits were mostly not significantly different from zero or favorable. To our knowledge, no genetic correlation between disease traits and traits selected in maternal rabbit lines have been reported previously. A study in a paternal rabbit line reported that genetic correlations between disease resistance traits and body weight at 63 or 70 days and carcass yield were not significantly different from zero or favorable (Gunia et al., 2015). However, in their review, Stear et al. (2001) observed that genetic correlations between production and disease resistance traits in livestock can be either favorable or unfavorable. The genetic correlations between production traits and Resist_S or Resist_Ch were very similar, despite different magnitudes of genetic variance for Resist in S and Ch.

Selection Strategies to Improve Disease Resistance Across Environments

Our aim is to reduce the prevalence of disease through selection for host resistance. All the breeding schemes tests showed that we could expect a reduction of disease incidence of 4 to 6% per generation (for the same level of exposure of rabbits to pathogens). If we assume linear genetic progress for the disease resistance traits, we could expect a reduction of prevalence of 41–31% for Resist_Ch and 26–21% for Resist_S over 5 generations. However, as stated by (Mackenzie and Bishop, 1999), predicting the consequences and benefits of selection is a difficult step,

because altering the genetics of individual animals affects the epidemiology of the disease at the population level. Indeed, if we select rabbits for resistance to disease, pathogen exposure is likely to change. As the number of susceptible animals in the population decreases, pathogen transmission among rabbits will also decrease. Bishop and Stear (1997) modeled the response to selection for resistance to parasites in sheep and obtained better responses to selection than predicted by quantitative genetic theory due to changes in the epidemiology of the disease.

The genetic progress on disease resistance traits was always favorable, even when disease resistance traits were not included in the breeding objective. This result is due to favorable correlations with the other traits. However, as some genetic correlation between production traits and Resist_S or Resist_Ch were not significantly different from zero, the “true” correlated response on disease resistance may be null if we select for HProduction. Except for the breeding scheme with HProduction, the genetic progress on Resist_S was very stable, probably because the amount of information available was always high. For this trait, records on the individual performances of the selection candidates as well as for a large number of sibs were always available for all scenarios including a disease resistance trait in the breeding objective. The genetic correlation between Resist_S and Resist_Ch therefore enabled good genetic progress for Resist_S even when this trait was not directly included in the breeding objective. This relatively high correlation between Resist_S and Resist_Ch and the higher heritability of Resist_Ch could explain why the scenario “Resist_Ch” gave the highest genetic gain for both Resist_Ch and Resist_S. Recording Resist_Ch always resulted in higher genetic progress for this trait, as could be expected. This finding is in accordance with previous results reported in the literature. Testing half-sibs under commercial conditions is considered as a good option to maintain genetic gain in the presence of G×E (Mulder and Bijma, 2005). Improving disease resistance in both environments is important, due to the high variability of environments on commercial farms (climate, pathogen loads). Some commercial farms apply high biosecurity measures and their environment is similar to the S environment, while on others the conditions are closer to our Ch environments. The breeding objectives presented here were based on the desired gain methodology, where the weight applied to the trait is based on the target genetic progress for the breeding company. The weight given to Resist in our simulations was high. Other scenarios with more balanced weights among traits may have led to less progress for Resist. Another methodology to define the breeding objective would be to consider economic weights and weighting the genetic gain in each environment by the relative importance of that environment (Mulder et al., 2006). These weights could reflect the size of the doe population in each environment. Economic weights have previously been derived for rabbit meat production (Cartuche et al., 2014). Litter size was the trait with the highest economic value, and was considered 5 times more important (expressed per standard deviation) than fattening survival. The discrepancy between the two methodologies can also reflect the strategic choices of different breeding companies.

CONCLUSION

Selection on non-specific disease resistance or tolerance using simple observations seems to be feasible. The trait is heritable, and the genetic correlations with the other traits under selection are not significantly different from zero or favorable. G×E interactions exist for this trait between selection and challenging environments. Therefore, recording this trait in both environments results in higher genetic progress. Long-term prediction of the genetic gain is difficult, due to the probable changes in disease epidemiology caused by selection. Quantitative genetics theory predicts a reduction of disease incidence by 4–6% per generation. The true genetic gain is likely to be greater than that predicted on the present study. Such selection could have a major impact on the reduction of antibiotic use and on the improvement of animal welfare. The biological mechanisms underlying non-specific disease resistance

are not fully understood yet. Beside classical immune parameters, the gut microbiota has recently emerged as a key regulator of immunity. Further studies using high throughput and deep phenotyping approaches of extreme animals with the highest estimated breeding values for disease resistance and disease sensitivity are needed to unravel the mechanisms in play.

AUTHOR CONTRIBUTIONS

MG performed the analyses, critically interpreted the data, and prepared the manuscript. HerG made substantial contributions to the analyses and interpretation of data. ID made substantial contributions to the analysis, critically interpreted the data, and reviewed the manuscript. JH and MM designed the experimental protocol and critically interpreted the data. HÉLG made substantial contributions to the interpretation of data and edition of the manuscript.

REFERENCES

- Bishop, S. C., and Stear, M. J. (1997). Modelling responses to selection for resistance to gastro-intestinal parasites in sheep. *Anim. Sci.* 64, 469–478. doi: 10.1017/S1357729800016088
- Bishop, S. C., and Woolliams, J. A. (2010). On the genetic interpretation of disease data. *PLoS ONE* 5:e8940. doi: 10.1371/journal.pone.0008940
- Brascamp, E. W. (1984). Selection indices with constraints. *Anim. Breed. Abstr.* 52, 645–654.
- Cartuche, L., Pascual, M., Gómez, E. A., and Blasco, A. (2014). Economic weights in rabbit meat production. *World Rabbit Sci.* 22, 165–177. doi: 10.4995/wrs.2014.1747
- David, I., Bouvier, F., Banville, M., Canario, L., Flatres-grall, L., Balmiss, E., et al. (2015). The direct-maternal genetic correlation has little impact on genetic evaluations. *J. Anim. Sci.* 93, 5639–5647. doi: 10.2527/jas.2015-9548
- Doeschl-Wilson, A. B., and Kyriazakis, I. (2012). Should we aim for genetic improvement in host resistance or tolerance to infectious pathogens? *Front. Genet.* 3:272. doi: 10.3389/fgene.2012.00272
- Eady, S. J., Garreau, H., and Gilmour, A. R. (2007). Heritability of resistance to bacterial infection in meat rabbits. *Livest. Sci.* 112, 90–98. doi: 10.1016/j.livsci.2007.01.158
- Garcia, M. L., and Baselga, M. (2002). Estimation of correlated response on growth traits to selection in litter size of rabbits using a cryopreserved control population and genetic trends. *Livest. Prod. Sci.* 78, 91–98. doi: 10.1016/S0301-6226(02)00093-3
- Garreau, H., Brad, S., Hurtaud, J., Guitton, E., Cauquil, L., Licois, D., et al. (2012). “Divergent selection for digestive disorders in two commercial rabbit lines: Response of crossbred young rabbits to an experimental inoculation of *Escherichia coli* O-103,” in *10th World Rabbit Congress*. (Sharm El-Sheikh), 153–157.
- Gianola, D. (1982). To estimate genetic correlations, we used analysis methods for continuous data, which are not theoretically optimal. The suitable methodology is the threshold model (Gianola, 1982). However, assumption of a continuous distribution for these traits is just. *J. Anim. Sci.* 54, 1079–1096. doi: 10.2527/jas1982.5451079x
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2009). *ASReml User Guide Release 3.0*. Hemel Hempstead: VSN International Ltd. Available online at: <https://www.vsnl.co.uk/downloads/asrem1/release3/UserGuide.pdf>
- Glass, E. J. (2012). The molecular pathways underlying host resistance and tolerance to pathogens. *Front. Genet.* 3:263. doi: 10.3389/fgene.2012.00263
- Gunia, M., David, I., Hurtaud, J., Maupin, M., Gilbert, H., and Garreau, H. (2015). Resistance to infectious diseases is a heritable trait in rabbits. *J. Anim. Sci.* 93, 5631–5638. doi: 10.2527/jas.2015-9377
- Guy, S. Z. Y., Thomson, P. C., and Hermes, S. (2012). Selection of pigs for improved coping with health and environmental challenges: breeding for resistance or tolerance? *Front. Genet.* 3:281. doi: 10.3389/fgene.2012.00281
- Hermesch, S., Li, L., Doeschl-Wilson, A. B., and Gilbert, H. (2015). Selection for productivity and robustness traits in pigs. *Anim. Prod. Sci.* 55, 1437–1447. doi: 10.1071/AN15275
- Herrero-Medrano, J. M., Mathur, P. K., ten Napel, J., Rashidi, H., Alexandri, P., Knol, E. F., et al. (2015). Estimation of genetic parameters and breeding values across different challenge environments to select for robust pigs. *J. Anim. Sci.* 93, 1494–1502. doi: 10.2527/jas.2014-8583
- Houle, D., and Meyer, K. (2015). Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *J. Evol. Biol.* 28, 1542–1549. doi: 10.1111/jeb.12674
- Kadarmideen, H. N., Thompson, R., Coffey, M. P., and Kossabati, M. A. (2003). Genetic parameters and evaluations from single- and multiple-trait analysis of dairy cow fertility and milk production. *Livest. Prod. Sci.* 81, 183–195. doi: 10.1016/S0301-6226(02)00274-9
- Kause, A., and Ødegård, J. (2012). The genetic analysis of tolerance to infections: a review. *Front. Genet.* 3:262. doi: 10.3389/fgene.2012.00262
- Lewis, C. R. G., Torremorell, M., Galina-Pantoja, L., and Bishop, S. C. (2008). Genetic parameters for performance traits in commercial sows estimated before and after an outbreak of porcine reproductive and respiratory syndrome. *J. Anim. Sci.* 87, 876–884. doi: 10.2527/jas.2008-0892
- Loussouarn, V., Robert, R., and Garreau, H. (2012). “Selection for weaning weight in Hyla breed: genetic parameters and trends,” in *10th World Rabbit Congress* (Sharm El-Sheikh), 189–193.
- Mackenzie, K., and Bishop, S. C. (1999). A discrete-time epidemiological model to quantify selection for disease resistance. *Anim. Sci.* 69, 543–551. doi: 10.1017/S1357729800051390
- Matos, C. A., Thomas, D. L., Gianola, D., Young, L. D., Matos, C. A. P., Thomas, D. L., et al. (1997b). Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models: II. Goodness of fit and predictive ability. *J. Anim. Sci.* 75, 88–94.
- Matos, C. A. P., Thomas, D. L., Gianola, D., Tempelman, R. J., and Young, L. D. (1997a). Genetic analysis of discrete reproductive traits in sheep using linear and nonlinear models. I. Estimation of genetic parameters. *J. Anim. Sci.* 75, 76–87.
- Meijering, A., and Gianola, D. (1985). Linear versus nonlinear methods of sire evaluation for categorical traits: a simulation study. *Genet. Sel. Evol.* 17, 115–132. doi: 10.1186/1297-9686-17-1-115
- Merk, J. W. M. (1989). Genotype × environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. *Livest. Prod. Sci.* 22, 325–339. doi: 10.1016/0301-6226(89)90064-X

- Merks, J. W. M., Mathur, P. K., and Knol, E. F. (2012). New phenotypes for new breeding goals in pigs. *Animal* 6, 535–543. doi: 10.1017/S1751731111002266
- Mocé, M. L., and Santacreu, M. A. (2010). "Genetic Improvement Of Litter Size In Rabbits," in *9th World Congress on Genetics Applied to Livestock Production*. (Leipzig), 121–129.
- Morrell, C. H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics* 54, 1560–1568. doi: 10.2307/2533680
- Mulder, H. A., and Bijma, P. (2005). Effects of genotype x environment interaction on genetic gain in breeding programs. *J. Anim. Sci.* 83, 49–61. doi: 10.2527/2005.83149x
- Mulder, H. A., Veerkamp, R. F., Ducro, B. J., van Arendonk, J. A. M., and Bijma, P. (2006). Optimization of dairy cattle breeding programs for different environments with genotype by environment interaction. *J. Dairy Sci.* 89, 1740–1752. doi: 10.3168/jds.S0022-0302(06)72242-1
- Phocas, F., Belloc, C., Bidanel, J., Delaby, L., Dourmad, J. Y., Dumont, B., et al. (2016). Review: towards the agroecological management of ruminants, pigs and poultry through the development of sustainable breeding programmes: I-selection goals and criteria. *Animal* 10, 1749–1759. doi: 10.1017/S1751731116000926
- Ragab, M., Ramon, J., Rafel, O., Quintanilla, R., Piles, M., and Sanchez, J. P. (2015). "Paramètres génétiques des phénotypes liés aux maladies chez le lapin en engraissement nourri avec deux regimes alimentaires differents," in *16ème Journées de la Recherche Cunicole* (Le Mans), 69–72. Available online at: <https://www.cuniculture.info/Docs/Magazine/Magazine2015/Fichiers-pdf-JRC/14-Ragab.pdf>
- Ramirez-Valverde, R., Misztal, I., and Bertrand, J. (2001). Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. *J. Anim. Sci.* 79, 333–338. doi: 10.2527/2001.792333x
- Riley, D. G., and Van Wyk, J. A. (2009). Genetic parameters for FAMACHA (c) score and related traits for host resistance/resilience and production at differing severities of worm challenge in a Merino flock in South Africa. *Vet. Parasitol.* 164, 44–52. doi: 10.1016/j.vetpar.2009.04.014
- Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics* 15, 469–485. doi: 10.2307/2527750
- Rothschild, M. F. (1998). "Rothschild, 1998 Selection for Diseases Resistance in pig," in *National Swine Improvement Federation Conference* Available online at: <http://www.nsif.com/conferences/1998/rothschild.htm>
- Rutten, M. J. M., Bijma, P., Woolliams, J. A., and van Arendonk, J. A. M. (2002). SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. *J. Hered.* 93, 456–458. doi: 10.1093/jhered/93.6.456
- Stear, M. J., Bishop, S. C., Mallard, B. A., and Raadsma, H. (2001). The sustainability, feasibility and desirability of breeding livestock for disease resistance. *Res. Vet. Sci.* 71, 1–7. doi: 10.1053/rvsc.2001.0496

Conflict of Interest Statement: This study was conducted under a research contract between INRA and Hypharm supervised by HerG. JH and MM are employed by Hypharm. However, the possible conflict of interest did not interfere with the outcome of this paper.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Gunia, David, Hurtaud, Maupin, Gilbert and Garreau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Genomic Architecture of Fowl Typhoid Resistance in Commercial Layers

Androniki Psifidi^{1,2*}, Kay M. Russell¹, Oswald Matika¹, Enrique Sánchez-Molano¹, Paul Wigley³, Janet E. Fulton⁴, Mark P. Stevens¹ and Mark S. Fife⁵

¹ The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, United Kingdom,

² Royal Veterinary College, University of London, Hatfield, United Kingdom, ³ Department of Infection Biology, Institute for Infection and Global Health, University of Liverpool, Neston, United Kingdom, ⁴ Hy-Line International, Dallas Center, IA, United States, ⁵ The Pirbright Institute, Surrey, United Kingdom

OPEN ACCESS

Edited by:

Martien Groenen,
Wageningen University and Research,
Netherlands

Reviewed by:

Xiangdong Ding,
China Agricultural University, China
Shaojun Liu,
Hunan Normal University, China

*Correspondence:

Androniki Psifidi
androniki.psifidi@roslin.ed.ac.uk;
apsifidi@rvc.ac.uk

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 30 May 2018

Accepted: 15 October 2018

Published: 19 November 2018

Citation:

Psifidi A, Russell KM, Matika O,
Sánchez-Molano E, Wigley P,
Fulton JE, Stevens MP and Fife MS
(2018) The Genomic Architecture
of Fowl Typhoid Resistance
in Commercial Layers.
Front. Genet. 9:519.
doi: 10.3389/fgene.2018.00519

Salmonella enterica serovar Gallinarum causes devastating outbreaks of fowl typhoid across the globe, especially in developing countries. With the use of antimicrobial agents being reduced due to legislation and the absence of licensed vaccines in some parts of the world, an attractive complementary control strategy is to breed chickens for increased resistance to *Salmonella*. The potential for genetic control of salmonellosis has been demonstrated by experimental challenge of inbred populations. Quantitative trait loci (QTL) associated with resistance have been identified in many genomic regions. A major QTL associated with systemic salmonellosis has been identified in a region termed SAL1. In the present study, two outbreaks of fowl typhoid in 2007 and 2012 in the United Kingdom were used to investigate the genetic architecture of *Salmonella* resistance in commercial laying hens. In the first outbreak 100 resistant and 150 susceptible layers were genotyped using 11 single nucleotide polymorphism (SNP) and 3 microsatellite markers located in the previously identified SAL1 region on chromosome 5. From the second outbreak 100 resistant and 200 susceptible layers, belonging to a different line, were genotyped with a high-density (600 K) genome-wide SNP array. Substantial heritability estimates were obtained in both populations ($h^2 = 0.22$ and 0.26 , for the layers in the first and second outbreak, respectively). Significant associations with three markers on chromosome 5 located close to *AKT1* and *SIVA1* genes, coding for RAC- α serine/threonine protein kinase, and the CD27-binding protein *SIVA1*, respectively, were identified in the first outbreak. From analysis of the second outbreak, eight genome-wide significant associations with *Salmonella* resistance were identified on chromosomes 1, 6, 7, 11, 23, 24, 26, 28 and several others with suggestive genome-wide significance were found. Pathway and network analysis revealed the presence of many innate immune pathways related to *Salmonella* resistance. Although, significant associations with SNPs located in the SAL1 locus were not identified by the genome-wide scan for layers from the second outbreak, pathway analysis revealed P13K/AKT signaling as the most significant pathway. In summary, resistance to fowl typhoid is a heritable polygenic trait that could possibly be enhanced through selective breeding.

Keywords: fowl typhoid, chicken, layers, disease outbreak, GWAS, pathway

INTRODUCTION

Salmonella enterica serovar Gallinarum causes a systemic bacterial disease mainly in adult poultry known as fowl typhoid. Outbreaks of this disease can have huge financial consequences with infected flocks having reduced egg production and a high percentage of mortality (Shivaprasad, 2000; Barrow and Freitas Neto, 2011). Regulations across the European Union compel poultry producers to control *Salmonella* in their layer and broiler breeder flocks. For example, in the United Kingdom, the Poultry Health Scheme routinely tests farms for the presence of *S. Gallinarum* resulting in rare occurrence of the disease after a prolonged control strategy (Poultry Health Scheme Handbook, 2013; Wigley, 2017). Despite such control measures, some outbreaks have been reported in recent years for both caged layers and backyard flocks in the United Kingdom indicating that outbreaks do still occur with devastating effects (Cobb et al., 2005; Parmar and Davies, 2007). More worrying, fowl typhoid has re-emerged in recent years in developing countries that have also established sanitary measures and official programs to prevent and control the disease. However, the disease remains endemic with cyclic or seasonal outbreaks related mainly to disease management (Revolledo, 2018). Therefore, a pressing need exists for complementary strategies to control the disease (Barbour et al., 2015; Guo et al., 2016; Celis-Estupinan et al., 2017; Pal et al., 2017; Weerasooriya et al., 2017).

Genetic selection for birds resistant to *S. Gallinarum* has been seen as an attractive solution for the control of fowl typhoid since the 1930's (Lambert and Knox, 1932). Inbred chicken lines have been described that exhibit heritable differences in resistance to systemic salmonellosis, including following oral *S. Gallinarum* inoculation or intravenous administration of *S. Typhimurium* (Bumstead and Barrow, 1993; Mariani et al., 2001). These lines have been extensively studied over the past 35 years, and crosses between these lines have been used to identify quantitative trait loci (QTL) for *Salmonella* resistance. A region on chromosome 5, termed *SAL1*, has been identified in multiple independent studies as having a protective role against systemic salmonellosis in the chicken (Mariani et al., 2001; Kaiser and Lamont, 2002; Tilquin et al., 2005; Calenge et al., 2010; Redmond et al., 2011). We refined the *SAL1* major QTL by mapping resistance in a 6th generation backcross with inbred lines 6I (resistant) and 15I (susceptible) using a high-density SNP panel (Fife et al., 2009). The refined *SAL1* region contains 14 genes with some noticeable candidates that have previously been linked with *Salmonella* resistance in other species, such as the RAC-alpha serine/ threonine protein kinase homolog, *AKT* (Fife et al., 2009). It is noteworthy that distinct QTL have been associated with enteric carriage of *S. Typhimurium* (Fife et al., 2011).

The present study builds on and extends our previous studies in inbred lines, aiming to dissect the genomic architecture of fowl typhoid resistance using two different United Kingdom commercial layer populations which suffered from natural outbreaks of fowl typhoid. We conducted variance component analyses to estimate genetic parameters and genomic association studies to identify genomic regions controlling fowl typhoid resistance. We also performed gene enrichment and pathway

analyses to identify candidate genes within the relevant genomic regions.

MATERIALS AND METHODS

Ethics Statement

All animal experiments were conducted in accordance with the revised Animals (Scientific Procedures) Act 1986 (project license PPL40/3652) with the approval of the local Ethical Review Body.

Study Population

Two different commercial laying hen populations suffering from two separate *S. Gallinarum* outbreaks of fowl typhoid, in 2007 and 2012 in the United Kingdom, were used in this study. From the first outbreak, blood and liver samples from 250 layers (150 susceptible and 100 resistant) were collected.

The second outbreak affected a layer farm with 375,000 birds. While most of the infected birds succumbed to infection, about 0.1% of the birds showed some level of resistance, with only mild clinical signs. Ultimately all remaining birds were culled on humane grounds, to prevent further spread of infection. From this outbreak, blood, spleen, and liver samples were collected from 300 layers (200 susceptible and 100 resistant). Three liver samples were collected from each bird, one in tissue storage reagent *RNAlater*[®], one in formalin for histological analysis, and one in phosphate-buffered saline (PBS) for enumeration of viable bacteria.

The collection of samples was performed by qualified veterinarians: samples were collected from birds raised in the same pens; live birds were culled and classified based on the observed pathology (lesions in liver, spleen, or ovary) as resistant or susceptible. Susceptible birds had extensive pathology implying potential death from lesions in the next 24 h. Resistant birds had no overt gross lesions on post mortem examination, with limited clinical signs.

For the first outbreak prevalence data was unavailable. For the second outbreak, the rate of infection varied between the six poultry houses on the affected premises. Levels of mortality consistent with clinical signs of fowl typhoid were recorded for the second outbreak with peak levels at approximately 3000 birds per day across the farm. Toward the end of the outbreak approximately 33% of birds had succumbed to infection. Birds for this study were sampled from the poultry house with the highest reported prevalence.

Phenotyping

For the first outbreak the trait was binary [0/1, case (susceptible)-control (resistant)]. For the second outbreak *S. Gallinarum* load in liver was determined in colony-forming units (CFU)/gram as described previously (Mariani et al., 2001). Briefly, liver samples in PBS were weighed and homogenized in an equal v/w of PBS. The homogenized liver tissue was serially diluted and plated on Modified Brilliant Green Agar (Oxoid, United Kingdom), incubated overnight, and the numbers of bacterial colonies were counted. The number of CFU/g was log transformed in order to

normalize the distribution. The trait for the second outbreak was analyzed both using continuous as well as binary phenotypes.

Histology and Assessment of Pathogenicity

Histological analyses were performed on liver and spleen samples from birds from the second outbreak. Samples of liver and spleen were fixed in formalin, paraffin-wax embedded then cut and stained with haematoxylin and eosin by the Department of Veterinary Pathology, University of Liverpool. Tissues were observed and analyzed blind as described previously (Parsons et al., 2013).

Assessment of pathogenicity of the strain isolated from the second outbreak in an experimental infection model was made in comparison with two well characterized *S. Gallinarum* isolates SG9 and 287/91 (Jones et al., 2001), as described previously (Langridge et al., 2015). Briefly, groups of five 3-week-old *Salmonella*-free commercial brown egg layer chickens (Lohmann Brown) were infected orally with 10^8 CFU of each of the *S. Gallinarum* isolates or remained as an uninfected control. At 6 days post challenge all birds were killed and at post mortem examination the spleen, liver, and caecal contents were removed for enumeration of viable *Salmonella* on selective Modified Brilliant Green Agar (Oxoid, United Kingdom) as detailed previously (Langridge et al., 2015).

Genotyping

All the birds from the first outbreak were genotyped using 11 custom-made SNP and 3 microsatellites markers located in the previously identified *SAL1* region on chromosome 5 (Fife et al., 2009). A full list of these markers is displayed in **Supplementary Table S1**. All the birds from the second outbreak were genotyped with the 600 K high density genome-wide SNP array (Affymetrix® Axiom® HD) (Kranis et al., 2013).

Heritability Analyses

Genetic parameters were estimated for *S. Gallinarum* resistance for the first and the second outbreak using a mixed linear univariate model that included the population principal components (for the second outbreak only) as a covariate effect, and the random effect of the individual bird. Genetic relationships between birds were calculated based on SNP genotypes using the genome-wide efficient mixed model association (GEMMA) algorithm (Zhou and Stephens, 2014) and included in the analyses. For the second outbreak the continuous phenotypes were used to estimate the variance components. The heritability of each trait was calculated as the ratio of the additive genetic to the total phenotypic variance. All above analyses were performed separately for each outbreak using the ASReml 4.0 software (Gilmour et al., 2009).

Genomic Association Analyses

Single-Marker Genomic Association Analyses

For the first outbreak a single marker association analysis where the SNP genotype was fitted as a fixed effect and the genomic

relatedness matrix was fitted as a random polygenic effect was performed using ASReml 4.0 software (Gilmour et al., 2009).

Data from the second outbreak were analyzed using two genome-wide association methodologies. Briefly, either a single SNP or a group of SNPs in sets of windows/ regions-using a regional heritability mapping approach (RHM)- were fitted as fixed effects.

The SNP genotype data were subjected to quality control measures using PLINK v1.09 (Purcell et al., 2007): minor allele frequency >0.05 , call rate $>95\%$ and Hardy-Weinberg equilibrium ($P > 10^{-6}$). After quality control, 297,560 SNP markers remained for further analysis. Positions of SNP markers were obtained using the Gal-gal5 assembly in Ensembl Genome Browser¹.

Population stratification was investigated using a genomic relatedness matrix generated from all individuals. This genomic relatedness matrix was converted to a distance matrix that was used to carry out classical multidimensional scaling analysis (MSA) using the GenABEL package of R (Aulchenko et al., 2007), to obtain its principal components.

The GEMMA algorithm (Zhou and Stephens, 2014) was used to perform GWAS analyses using a standard univariate linear mixed model in which the first four principal components were fitted as covariate effects to adjust for population structure and the genomic relatedness matrix among individuals was fitted as a polygenic effect. After Bonferroni correction for multiple testing, significance thresholds were $P \leq 1.68 \times 10^{-7}$ and $P \leq 3.36 \times 10^{-6}$ for genome-wide significant ($P \leq 0.05$) and suggestive (namely one false positive per genome scan) levels, respectively, corresponding to $-\log_{10}(P)$ of 6.77 and 5.47. The Chi-square (χ^2) test was implemented to validate the GWAS results. A P -value for each comparison (expected vs. observed values) was estimated based on the χ^2 statistics value for two degrees of freedom. The significance threshold was set at $P \leq 0.05$. The extent of linkage disequilibrium (LD) between significant SNPs located on the same chromosome regions was calculated using the r -square statistic of PLINK v1.09 (Purcell et al., 2007).

Regional Heritability Mapping

The RHM approach was used to analyse data from the second outbreak fitting genomic regions of 20 SNPs in sliding “windows” along each chromosome. RHM analyses were performed using the DISSECT software (Canela-Xandri et al., 2015) fitting the same fixed effects as the ones used in the single SNP GWAS described above. The significance of genomic regions was assessed with the likelihood ratio test statistic, which was used to compare the RHM model where both the whole genome and a genomic region were fitted as random effects against the base model that excluded the latter effect. A total of 14,878 regions were tested across the genome. After the adjustment, using Bonferroni correction, for multiple testing significance thresholds were $P \leq 3.37 \times 10^{-6}$ and $P \leq 6.72 \times 10^{-5}$ for genome-wide ($P \leq 0.05$) and suggestive (namely one false

¹www.ensembl.org

positive per genome scan) levels, respectively, corresponding to $-\log_{10}(P)$ of 5.47 and 4.17.

SNP and Candidate Region Annotation

All significant SNPs identified in the GWAS for the second *S. Gallinarum* outbreak were mapped to the reference genome and annotated by using the variant effect predictor² tool within the Ensembl database and the Gal-gal5 assembly. Moreover, the genes that were located 100 kb upstream and downstream of the significant SNPs were also annotated using the BioMart data mining tool³ and the Gal-gal5 assembly. We chose these 200 kb windows based on the average LD in commercial populations (less than 1 cM on average; Andreescu et al., 2007) and the fact that the chicken genome contains 250 kb per cM on average (International Chicken Genome Sequencing Consortium, 2004). This allowed us to catalog all the genes that were located in the vicinity of the identified significant SNPs and to create gene lists that contained the genes in the vicinity of all the significant SNPs identified for fowl typhoid resistance.

Pathway, Network and Functional Enrichment Analyses

Identification of potential canonical pathways and networks underlying the candidate genomic regions associated with resistance to the second *S. Gallinarum* outbreak was performed using the Ingenuity Pathway Analysis (IPA) program⁴. IPA constructs multiple possible upstream regulators, pathways, and networks that serve as hypotheses for the biological mechanism underlying the phenotypes based on a large-scale causal network derived from the Ingenuity Knowledge Base. Then, IPA infers the most suitable pathways and networks based on their statistical significance, after correcting for a baseline threshold (Krämer et al., 2014). The IPA score in the constructed networks can be used to rank these networks based on the *P*-values obtained using Fisher's exact test [IPA score or *P*-score = $-\log_{10}(P\text{-value})$].

The gene list for *S. Gallinarum* resistance was also analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; Dennis et al., 2003). In order to understand the biological meaning behind these genes, gene ontology (GO) was determined and functional annotation clustering analysis was performed. The *Gallus gallus* background information is available in DAVID and was used for the analysis. The enrichment score (ES) of the DAVID package is a modified Fisher exact *P*-value calculated by the software, with higher ES reflecting more enriched clusters. An ES greater than 1 means that the functional category is overrepresented.

RESULTS

Descriptive Statistics of Phenotypes

A mean three-log difference of liver *S. Gallinarum* viable counts between the resistant (average: 4.4 \log_{10} CFU/gr, standard

deviation: 1.66) and the susceptible (average: 7.4 \log_{10} CFU/gr, standard deviation: 0.77) birds from the second outbreak was detected, consistent with the pathology results. The maximum of liver count measured was 8.45 \log_{10} CFU/gr, while in 34 samples no viable *S. Gallinarum* was detected (minimum).

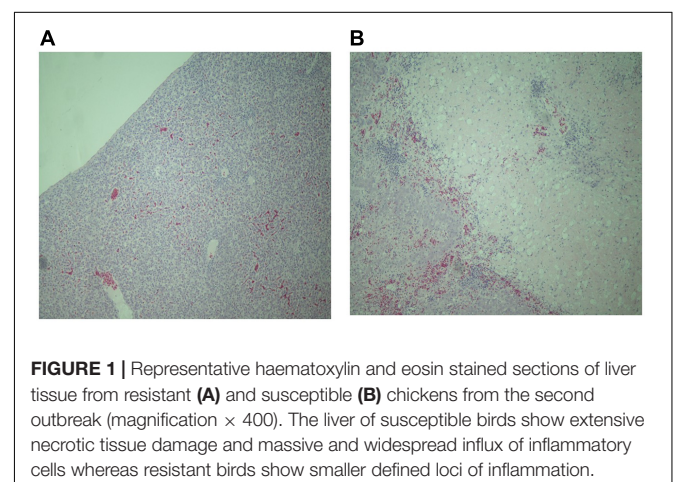
Histology and Assessment of Pathogenicity

As many samples were autolysed or degraded detailed scoring was not possible. However, analysis of tissues from six resistant and nine susceptible birds where the sample was not compromised, showed patterns of pathology similar with the ones previously described following experimental infection of resistant and susceptible inbred lines with *S. Gallinarum* (Wigley et al., 2002). Resistant birds showed signs of inflammation, largely restricted to specific foci in the liver (**Figure 1A**) and general inflammation in the spleen. In contrast susceptible birds showed greater levels of inflammation and large areas of necrotic damage in the liver (**Figure 1B**), with a high degree of inflammatory cell influx into the spleen with thickening of the splenic capsule and some areas of necrosis. These findings are consistent with observations in inbred lines exhibiting differential resistance following experimental infection (Mariani et al., 2001).

In experimental infection studies, a clonal isolate from the second outbreak was recovered in equivalent or greater numbers from the spleen and liver of orally challenged birds than 287/91 or SG9 (**Supplementary Figure S1**). This fulfills Koch's postulates and the outbreak strain may be considered typical of other *S. Gallinarum* strains in the pathology it elicits. None of the isolates were detected in the caecal contents at the time of post mortem examination.

Single-Marker Genomic Association Studies

Similar moderate heritability estimates for *S. Gallinarum* resistance were derived for both layer populations in the first ($h^2 = 0.22 \pm 0.01$) and second ($h^2 = 0.26 \pm 0.14$) outbreaks.



²<http://www.ensembl.org/Tools/VEP>

³<http://www.ensembl.org/biomart/martview/>

⁴www.ingenuity.com

Seven markers located in the *SAL1* locus on chromosome 5 were found to have a significant ($P < 0.05$) association with *S. Gallinarum* resistance in the layer population affected by the first outbreak. Details of the significant markers identified are presented in **Table 1**.

Multidimensional scaling analysis revealed four substructure principal components in the layer population affected by the second outbreak, which were subsequently included in the GWAS model to correct results for population stratification.

GWAS analysis identified six SNP markers genome-wide significantly associated with the log-transformed liver load of *S. Gallinarum* in layers from the second outbreak on chromosomes 1, 11, 23, 24, and 26 (P -values 7.36×10^{-10} to 1.63×10^{-7}) (**Table 2**). Additionally, 14 SNPs crossing the suggestive genome-wide significant threshold were identified on chromosomes 1, 2, 4, 6, 13, 19, 24, and 28 (**Table 2**). The Manhattan plot and the Q-Q plot for the GWAS results are displayed in **Figures 2A,B**.

The same significant associations on chromosomes 1, 23, 26, and 28 were identified by the GWAS analysis when the data was re-analyzed as a binary (case-control) trait (**Table 2**), although the ranking of the SNPs based on the P -values were different. With the case-control analysis the association on chromosome 28 attained genome-wide significance (P -values 2.41×10^{-12}). This approach identified also two new genome-wide significant associations on chromosomes 6 and 7 (P -values 8.81×10^{-9} to 1.08×10^{-8}) and new suggestive associations with markers on chromosomes 1, 3, 10, 11, and 23 (**Table 2**). All the significant associations identified by the GWAS were also found to be significant ($P < 0.05$) in the chi-square analysis. The Manhattan plot and the Q-Q plot for the GWAS results from the case-control analysis are displayed in **Figures 3A,B**. Significant SNPs that were located on the same chromosome were not in LD with the exception of the markers located on chromosome 13 ($r^2 > 0.90$).

Regional Heritability Mapping

The RHM mapping confirmed the significant associations on chromosomes 1, 11, 23, 24, and 26 previously identified by the GWAS (**Supplementary Table S2**). Moreover, RHM detected two more suggestive significant associations on chromosomes 2 and 11. Details of the significant SNP windows are presented in

TABLE 1 | List of SNPs associated with fowl typhoid resistance in the layer population from the first outbreak.

| SNP Name | Chromosome | Position | P-value |
|---------------|------------|----------|---------|
| SNP7 | 5 | 50401216 | 0.005 |
| SNP94 | 5 | 50471836 | 0.002 |
| SNP215 | 5 | 51415477 | 0.026 |
| SNP197 | 5 | 51685833 | 0.006 |
| SNP200 | 5 | 51686240 | <0.001 |
| SNP201 | 5 | 51686379 | <0.001 |
| SNP202 | 5 | 51739862 | 0.047 |

SNP markers in bold are spanning the *AKT1* gene.

Supplementary Table S2. The Manhattan plot and the Q-Q plot for the RHM results analysis are displayed in **Supplementary Table S2**.

SNP and Candidate Region Annotation

All of the significant markers identified for the first outbreak were located in intronic, intergenic, upstream and downstream gene regions with the exception of one SNP (SNP7) which corresponds to a missense variant within the Creatine Kinase B (*CKB*) gene. Four of the significant SNP markers spanned the *AKT1* gene. All these SNPs are intronic variants for the *AKT1* and also upstream and downstream variants for one microRNA (gga-mir-1771). The candidate region for fowl typhoid on chromosome 5 contained 16 protein coding genes and 2 microRNAs (**Supplementary Table S3**).

Most of the significant SNPs identified by the GWAS analyses for the second outbreak were located in intronic (34%), intergenic (24%) and upstream and downstream gene (14%) regions. However, four of the SNPs were localized in exonic regions. Specifically, Affx-51116866 corresponds to a missense variant within the Cell Adhesion Molecule 1 (*CADM1*) gene; Affx-51148005 corresponds to a missense variant within the TATA-Box Binding Protein Associated Factor 8 (*TAF8*) gene; Affx-51686897 corresponds to a missense variant within the AT-Rich Interaction Domain 5B (*ARID5B*) gene; Affx-51177949 corresponds to a synonymous variant within the Growth Differentiation Factor 3 (*GDF3*) gene. The above mentioned missense variants had a predicted moderate impact.

Most of the candidate regions for fowl typhoid resistance identified from the second outbreak contained multiple genes. In total 116 protein-coding genes and 4 microRNAs identified across the QTL regions for the second outbreak (**Supplementary Table S3**).

Pathway, Network and Functional Enrichment Analyses

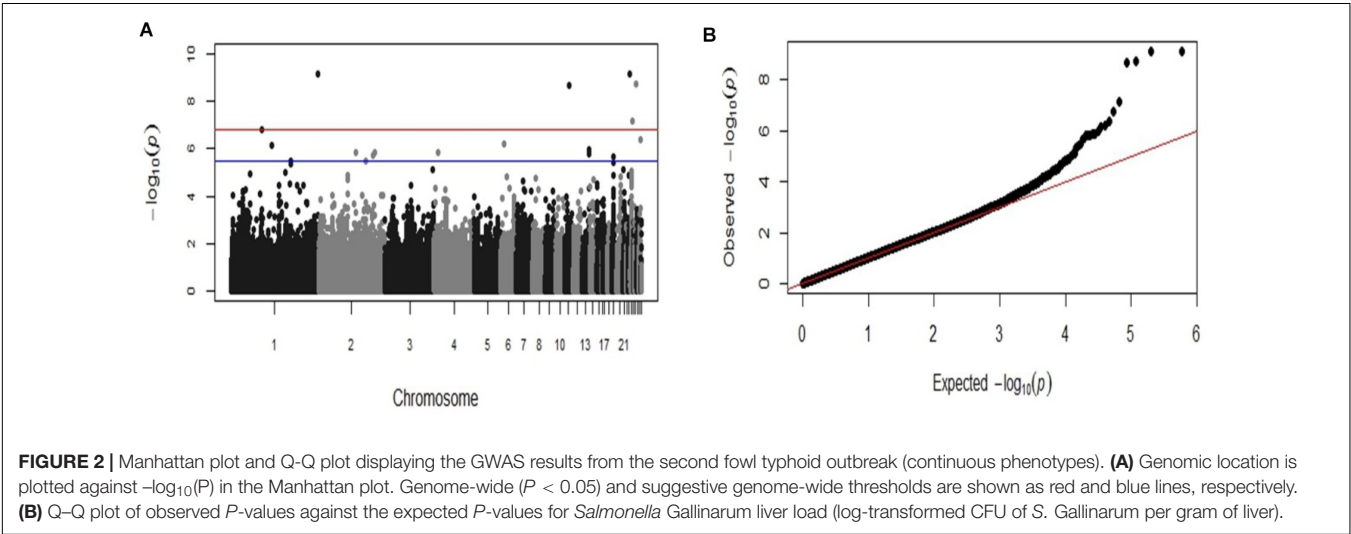
We reasoned that the corresponding QTL regions may contain genes contributing to a common pathway associated with *S. Gallinarum* resistance. We therefore identified the sets of annotated genes lying within the QTL intervals identified for the second outbreak and sought evidence of gene set enrichment. These genes were enriched for pathways involved in immune responses, both innate and adaptive, and cell-cycle regulation (**Figure 4**). The most enriched pathway was related to the PI3K/AKT signaling. Moreover, three networks of molecular interactions related to cell death and survival, and cell cycle, humoral immune response, hematological system development and function, and hematopoiesis were constructed using the list of genes in the candidate regions (**Figure 5**).

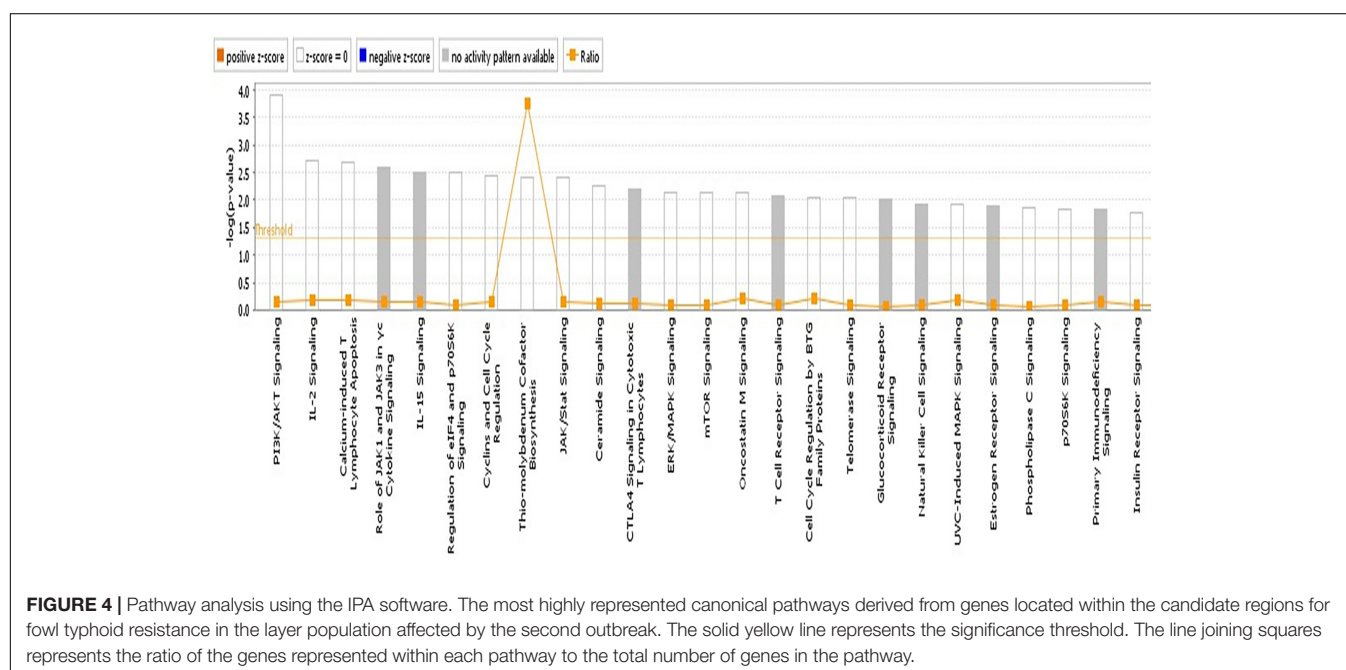
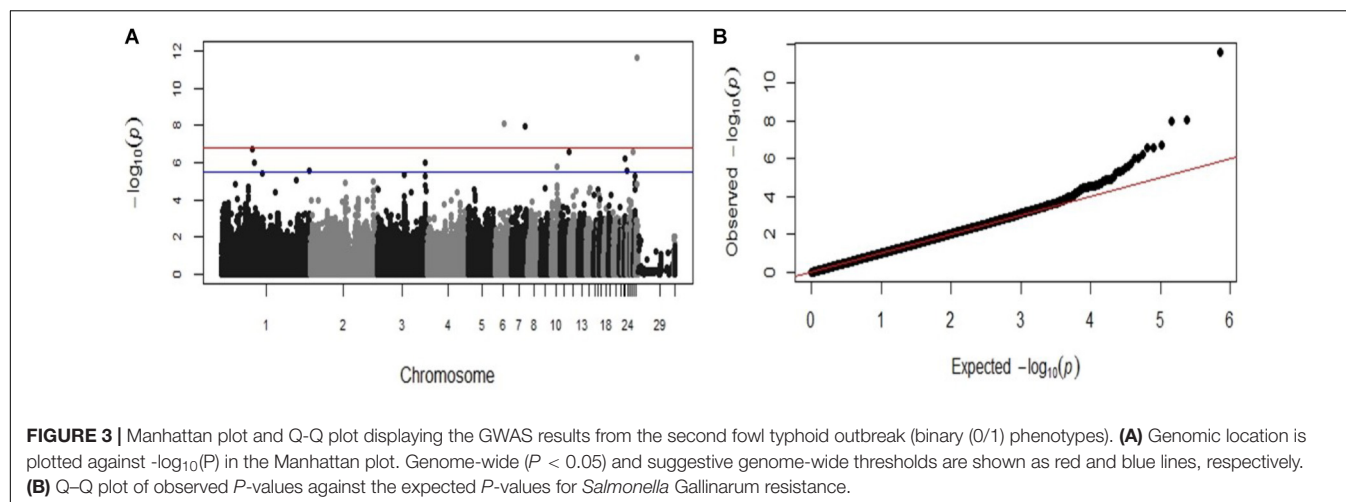
Functional annotation clustering analysis revealed the presence of enriched gene clusters related to protein kinase binding (E.S = 2.35, genes in the cluster: *PRKRIP1*, *CCND3*, *YWHAG*, *HDAC1*), positive regulation of immune system processes (E.S = 1.7, genes in the cluster: *CD247*, *SH2B2*, *CADM1*, *HPX*, *LCK*), hematopoiesis and immune system

TABLE 2 | List of SNPs associated with fowl typhoid resistance in the layers from the second outbreak.

| Phenotype | SNP name | Chr | Position | P-value |
|-----------|----------------------|-----------|------------------|------------------|
| Continues | Affx-50313880 | 1 | 194531778 | 7.368E-10 |
| | Affx-51098463 | 23 | 5081233 | 7.468E-10 |
| | Affx-51148005 | 26 | 5073473 | 1.852E-09 |
| | Affx-50516977 | 11 | 11072620 | 2.042E-09 |
| | Affx-51116866 | 24 | 4720171 | 7.372E-08 |
| | Affx-50405629 | 1 | 67549126 | 1.636E-07 |
| | Affx-51177949 | 28 | 3758677 | 4.249E-07 |
| | Affx-51686897 | 6 | 8299315 | 6.298E-07 |
| | Affx-50447114 | 1 | 91550805 | 6.893E-07 |
| | Affx-50617622 | 13 | 16337564 | 1.082E-06 |
| | Affx-51370634 | 4 | 10316583 | 1.373E-06 |
| | Affx-50841906 | 2 | 127840521 | 1.379E-06 |
| | Affx-50988352 | 2 | 85262253 | 1.468E-06 |
| | Affx-50617378 | 13 | 16238356 | 1.572E-06 |
| | Affx-50617564 | 13 | 16313839 | 1.572E-06 |
| | Affx-50832761 | 2 | 122511846 | 1.865E-06 |
| | Affx-50780736 | 19 | 4130929 | 2.219E-06 |
| | Affx-50193882 | 1 | 133009585 | 3.339E-06 |
| | Affx-50808404 | 2 | 107286660 | 3.424E-06 |
| | Affx-51107231 | 24 | 2183253 | 3.424E-06 |
| Binary | Affx-51177949 | 28 | 3758677 | 2.414E-12 |
| | Affx-51643081 | 6 | 20430378 | 8.818E-09 |
| | Affx-51739265 | 7 | 31474122 | 1.083E-08 |
| | Affx-50405629 | 1 | 67549126 | 2.015E-07 |
| | Affx-50538456 | 11 | 19872676 | 2.724E-07 |
| | Affx-51148005 | 26 | 5073473 | 2.808E-07 |
| | Affx-51088276 | 23 | 2561442 | 6.146E-07 |
| | Affx-50414020 | 1 | 71999883 | 9.84E-07 |
| | Affx-51197199 | 3 | 107050735 | 1.026E-06 |
| | Affx-50476276 | 10 | 14891444 | 1.75E-06 |
| | Affx-51098463 | 23 | 5081233 | 2.67E-06 |

P-value from genomic association study (genome-wide significant in bold, suggestive significance otherwise).





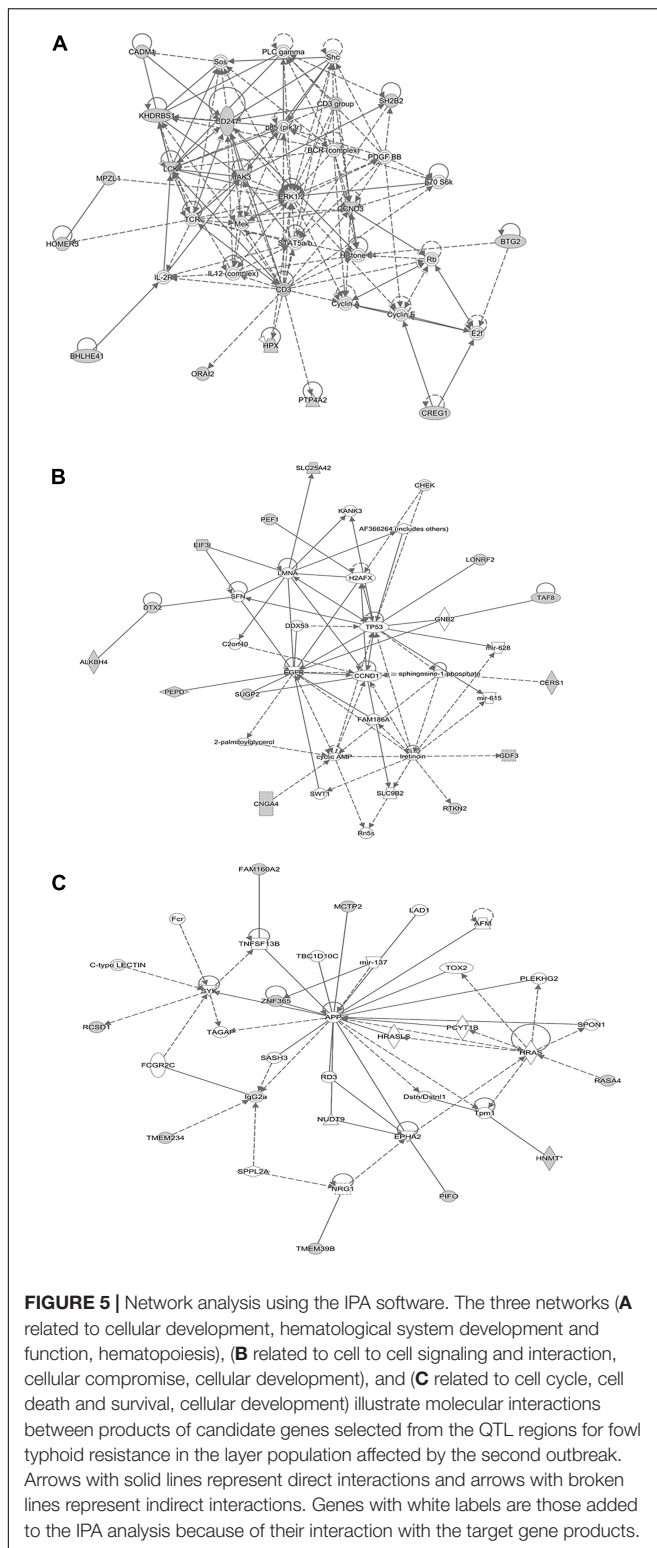
development (E.S = 1.1, genes in the cluster: *CEBPA*, *CEBPG*, *LCK*, *RTKN2*).

DISCUSSION

Our study set out to investigate the genetic basis of fowl typhoid resistance in commercial layers. Using samples from two natural disease outbreaks, we detected heritable genetic variation and identified genomic regions associated with resistance to the disease in two different layer populations. Putative candidate genes, canonical pathways and networks involved in the underlying molecular mechanisms of fowl typhoid resistance were also identified.

In terms of phenotype, there was on average a 3 \log_{10} difference in the recovery of viable *S. Gallinarum* between

resistant and susceptible birds from the second outbreak and differences in pathology that are consistent with those observed following experimental infection of inbred lines that exhibit heritable differences in resistance following oral *S. Gallinarum* or intravenous *S. Typhimurium* inoculation (Bumstead and Barrow, 1993; Mariani et al., 2001). Although much of the QTL-based mapping of the *SAL1* locus in inbred lines used intravenous infection of day old chicks with *S. Typhimurium*, the phenotype of resistance to experimental fowl typhoid is strongly expressed in older birds (Bumstead and Barrow, 1993) with quantitative differences of 3–4 \log_{10} CFU per gram of liver tissue between resistant and susceptible lines found 8 days after oral challenge with *S. Gallinarum* in 3-week-old birds (Wigley et al., 2002). Therefore, the pathology, phenotyping and histological results of the present study conducted in commercial layers are consistent with previous findings in inbred lines for fowl typhoid infection.



In addition, the present study provided further evidence for the role of *SAL1* locus in *Salmonella* resistance. *AKT1* is a promising candidate gene of this QTL region as the protein is known to be activated by *Salmonella* and to promote intracellular

net replication of the bacteria in mammalian cells (Steele-Mortimer et al., 2000; Kuijl et al., 2007). In the first outbreak the markers with the most significant association with the fowl typhoid spanned the *AKT1* gene. Although, significant associations with SNPs located in the *SAL1* locus were not identified by the genome-wide scan for layers from the second outbreak, pathway analysis revealed that the P13K/AKT signaling as the most significant pathway, implying that *AKT* pathway might play a role in *Salmonella* resistance. It is possible that other genes that are part of the P13/AKT pathway such as *JAK3*, *KRAS*, *GYS2*, *PPP2CA*, *YWHAG* might contribute to fowl typhoid resistance in the layers of the second outbreak since they belong to a different selection line and SNP markers proximal to these genes were identified in the GWAS analysis. Therefore, although the underlying mechanism might be similar, the causative mutation(s) might be different in the two populations. In addition, the phase of LD between the SNP markers and the causative mutation(s) might be different in the two different layer populations. *AKT* is a serine/threonine kinase that modulates multiple processes, in particular apoptosis, cell proliferation, and development (Hers et al., 2011). Depending on the cell type and stage of infection, apoptosis may play both positive and negative roles in control of *Salmonella* infection (Fink and Cookson, 2007). Nevertheless, the involvement of the other striking candidate gene, the CD27-binding protein *SIVA1*, in fowl typhoid resistance could not be excluded since the two candidate genes are in close proximity and significant markers were detected on either sides of these genes. *SIVA1* is a pro-apoptotic factor that induces cell death via a caspase-dependent pathway in human and murine cells (Prasad et al., 1997; Py et al., 2004). It has been also proposed that differences in the expression or function of *SIVA1* in the progeny of advanced inter-cross chicken lines may explain differences in the ability of heterophils from such birds to release heterophil extracellular traps via an apoptosis-like pathway (Redmond et al., 2011).

This is the first study, to our knowledge, that aimed to dissect the genetic architecture of fowl typhoid resistance using data of natural disease outbreaks. However, there are many previous genetic studies of systemic salmonellosis, *Salmonella* enteric carriage, carrier-state and antibody responses based on challenge experiments of *S. Enteritidis* and *S. Typhimurium* in crosses of inbred, and crosses of inbred with commercial chicken lines. Interestingly, many of the previously identified QTLs are overlapping or are in close proximity with the ones identified in the present study. The two QTLs we identified on chromosome 1 at position 67.5 and 91.5 Mb are closely located; the former with one identified in inbred lines for cloacal bacterial burden after oral challenge with *S. Enteritidis* (Tilquin et al., 2005) and the latter with one identified in broiler crosses for spleen bacterial burden after intra-oesophageal challenge (Kaiser and Lamont, 2002) and for vaccine response after subcutaneous challenge with *S. Enteritidis* (Kaiser et al., 2002). The QTLs on chromosome 1 (194.5 Mb) and chromosome 11 (20 Mb) overlap with QTLs found in inbred line crosses for carrier *Salmonella* state after oral challenge with *S. Enteritidis* (Calenge et al., 2011). Likewise, the QTL regions on chromosome 2

(122 Mb) and 4 overlap with previously identified QTLs for spleen bacterial burden after challenge with *S. Enteritidis* intra-oesophageal (Malek et al., 2004). The QTL on chromosome 3 overlaps with a QTL identified in advanced intercrosses of inbred lines with broilers for spleen bacterial burden after intra-oesophageal infection with *S. Enteritidis* (Hasenstein and Lamont, 2007). In the latter study, the gallinacin group of genes were considered good candidate genes for *Salmonella* resistance. The gallinacin-8 precursor (*AvBD8*) gene is also in close proximity with the significant marker identified on chromosome 3 in the present study. However, more studies are needed to confirm if this is the actual causative gene for this QTL. The QTLs that we identified on chromosomes 7, 19, 23, 24, and 26 are co-localized with previously identified QTLs in inbred chicken line crosses for *S. Enteritidis* caeca colonization after oral inoculation (Thanh-Son et al., 2012). Many immune genes (such as *LAT2/NTAL*, *TRAF3IP3*, *IRF6*) located within these QTL regions have been suggested as good candidate genes for *Salmonella* resistance.

The present study implemented a much higher density genome-wide genotyping platform compared to all the previous ones and was able to identify some novel QTLs. Moreover, two different approaches, GWAS and RHM, were implemented to further facilitate the QTL discovery. GWAS performs single marker analyses while RHM fits genomic regions of multiple SNPs as a single measure. Therefore, RHM has greater power compared to GWAS to identify loci where several alleles with small effects segregating. In addition, we implemented two different GWAS models, one using binary phenotypes and the other the continuous phenotypes. We used the binary phenotypes to be consistent with the phenotypes used to analyse the first outbreak, and the continuous ones to increase further the power of the study and overcome putative errors derived from misclassifications of cases and controls. The marker on chromosome 28 found to have the most significant association with fowl typhoid resistance, when the trait was analyzed as binary, is surrounded by many putative good candidate genes. Such genes related with immune response are the tyrosine-protein Janus kinase 3 (*JAK3*), the CREB regulator transcription coactivator 1 (*CRTC1*) and the cytokine receptor like factor 1 (*CRLF1*). The IPA analysis identified two canonical pathways related with JAK signaling among the most enriched pathways in this dataset: the JAK1 and JAK3 in the γ c cytokine regulation signaling and the JAK-Stat signaling. In addition, the immune related network with the highest IPA score had as one of the central molecules the *JAK3* protein. The JAK signaling family of tyrosine kinases are involved in cytokine receptor-mediated intracellular signal transduction. Specifically, *JAK3* mediates essential signaling events in both innate and adaptive immunity and plays a crucial role in hematopoiesis during T-cells development (Yamaoka et al., 2004). Multiple markers on chromosome 13 were found to have a significant association with fowl typhoid resistance. These markers span the follistatin-related protein 4 precursor (*FSTL4*) gene which is related with calcium metabolism and transportation. However, in close proximity (<0.5 Mb), immune genes of interest such as the Interleukin 3 precursor (*IL-3*), Interleukin 5 precursor (*IL5*),

and the Interferon Regulatory Factor 1 (*IRF1*), are located. The protein encoded by *IRF1* gene is a transcriptional regulator and tumor suppressor, serving as an activator of genes involved in both innate and acquired immune responses. The encoded protein activates the transcription of genes involved in the body's response to viruses and bacteria, playing a role in cell proliferation, apoptosis, immune and DNA damage response (Yoshida and Azuma, 1992; Taniguchi et al., 1995). In addition, it is involved in the regulation of interferon (*IFN*) and *IFN*-inducible genes that have been reported to be involved in host resistance to *Salmonella* infection (Thanh-Son et al., 2012).

CONCLUSION

We confirmed that resistance to fowl typhoid is a heritable complex polygenic trait. Co-localisation of many of the QTLs identified for fowl resistance with previous ones identified for systemic and enteric salmonellosis, and antibody responses implying that common underlying mechanisms of resistance to different *Salmonella* serovars segregating across chicken populations. These findings strengthen the interest of these regions for more refined analyses. According to our results breeding for enhanced fowl typhoid resistance in layers is possible. Although genomic selection is a valid approach to enhance disease resistance in chickens, as has been reported previously (Legarra et al., 2011), identification of the causative genes and mutations could expedite selection through different weighting of the validated selectable markers or precision breeding. However, further studies are needed to identify the causative genes and mutations.

DATA AVAILABILITY STATEMENT

All data are available at Figshare (doi: 10.6084/m9.figshare.7205702).

AUTHOR CONTRIBUTIONS

MF, MS conceived and designed the genetic studies of fowl typhoid and secured funding. AP and OM performed the genetic parameter analysis, collated and edited the genotyping data, and performed the single marker genomic analysis. AP and ES-M performed the regional heritability mapping. AP performed the pathway analysis and wrote the manuscript with input from KR. PW performed the histological analysis and experimental infections. MF, MS, JF, KR, and AP interpreted these results. All other co-authors provided manuscript editing and feedback. All authors read and approved the final manuscript.

FUNDING

The authors gratefully acknowledge funding from the Biotechnology & Biological Sciences Research Council (BBSRC)

and Hy-Line International (BB/J015296/1) and BBSRC strategic investment at The Pirbright Institute (BB/J016837/1 and BB/P013740/1) and The Roslin Institute (BB/J004227/1 and BBS/E/D/20002172).

ACKNOWLEDGMENTS

We would like to extend our gratitude and thanks for the help of the vets collecting the samples and performing the clinical

analyses, and to the farmers who allowed us to use these samples for research purposes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00519/full#supplementary-material>

REFERENCES

- Andreescu, C., Avendano, S., Brown, S. R., Hassen, A., Lamont, S. J., and Dekkers, J. C. (2007). Linkage disequilibrium in related breeding lines of chickens. *Genetics* 177, 2161–2169. doi: 10.1534/genetics.107.082206
- Aulchenko, Y. S., Ripke, S., Isaacs, A., and Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi: 10.1093/bioinformatics/btm108
- Barbour, E. K., Ayyash, D. B., Alturkistni, W., Alyahiby, A., Yaghoor, S., Iyer, A., et al. (2015). Impact of sporadic reporting of poultry *Salmonella* serovars from selected developing countries. *J. Infect. Dev. Ctries.* 9, 1–7. doi: 10.3855/jidc.5065
- Barrow, P. A., and Freitas Neto, O. C. (2011). Pullorum disease and fowl typhoid—new thoughts on old diseases: a review. *Avian Pathol.* 40, 1–13. doi: 10.1080/03079457.2010.542575
- Bumstead, N., and Barrow, P. (1993). Resistance to *Salmonella* gallinarum, *S. pullorum*, and *S. enteritidis* in inbred lines of chickens. *Avian Dis.* 37, 189–193. doi: 10.2307/1591473
- Calenge, F., Kaiser, P., Vignal, A., and Beaumont, C. (2010). Genetic control of resistance to salmonellosis and to *Salmonella* carrier-state in fowl: a review. *Genet. Sel. Evol.* 42:11. doi: 10.1186/1297-9686-42-11
- Calenge, F., Vignal, A., Demars, J., Feve, K., Menanteau, P., Velge, P., et al. (2011). New QTL for resistance to *Salmonella* carrier-state identified on fowl microchromosomes. *Mol. Genet. Genomics* 285, 237–243. doi: 10.1007/s00438-011-0600-9
- Canela-Xandri, O., Law, A., Gray, A., Woolliams, J. A., and Tenesa, A. (2015). A new tool called DISSECT for analysing large genomic data sets using a big data approach. *Nat. Commun.* 6:10162. doi: 10.1038/ncomms10162
- Celis-Estupinan, A., Batista, D. F. A., Cardozo, M. V., Secundo De Souza, A. I., Rodrigues Alves, L. B., Maria De Almeida, A., et al. (2017). Further investigations on the epidemiology of fowl typhoid in Brazil. *Avian Pathol.* 46, 416–425. doi: 10.1080/03079457.2017.1299922
- Cobb, S. P., McVicar, C. M., Davies, R. H., and Ainsworth, H. (2005). Fowl typhoid in caged layer birds. *Vet. Rec.* 157:268. doi: 10.1136/vr.157.9.268
- Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:R60. doi: 10.1186/gb-2003-4-9-r60
- Fife, M. S., Howell, J. S., Salmon, N., Hocking, P. M., Van Diemen, P. M., Jones, M. A., et al. (2011). Genome-wide SNP analysis identifies major QTL for *Salmonella* colonization in the chicken. *Anim. Genet.* 42, 134–140. doi: 10.1111/j.1365-2052.2010.02090.x
- Fife, M. S., Salmon, N., Hocking, P. M., and Kaiser, P. (2009). Fine mapping of the chicken salmonellosis resistance locus (SAL1). *Anim. Genet.* 40, 871–877. doi: 10.1111/j.1365-2052.2009.01930.x
- Fink, S. L., and Cookson, B. T. (2007). Pyroptosis and host cell death responses during *Salmonella* infection. *Cell. Microbiol.* 9, 2562–2570. doi: 10.1111/j.1462-5822.2007.01036.x
- Gilmour, A. R., Cullis, B. R., and Thompson, R. (2009). *ASREML User Guide, Release 3.0*. Orange, NSW: NSW Department of Primary Industries.
- Guo, R., Geng, S., Jiao, H., Pan, Z., Chen, X., and Jiao, X. (2016). Evaluation of protective efficacy of a novel inactivated *Salmonella* Pullorum ghost vaccine against virulent challenge in chickens. *Vet. Immunol. Immunopathol.* 173, 27–33. doi: 10.1016/j.vetimm.2016.03.015
- Hasenstein, J. R., and Lamont, S. J. (2007). Chicken gallinacin gene cluster associated with *Salmonella* response in advanced intercross line. *Avian Dis.* 51, 561–567. doi: 10.1637/0005-2086(2007)51[561:CGGCAW]2.0.CO;2
- Hers, I., Vincent, E. E., and Tavaré, J. M. (2011). Akt signalling in health and disease. *Cell. Signal.* 23, 1515–1527. doi: 10.1016/j.cellsig.2011.05.004
- International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.
- Jones, M. A., Wigley, P., Page, K. L., Hulme, S. D., and Barrow, P. A. (2001). *Salmonella enterica* serovar Gallinarum requires the *Salmonella* pathogenicity island 2 type III secretion system but not the *Salmonella* pathogenicity island 1 type III secretion system for virulence in chickens. *Infect. Immun.* 69, 5471–5476. doi: 10.1128/IAI.69.9.5471-5476.2001
- Kaiser, M. G., Deeb, N., and Lamont, S. J. (2002). Microsatellite markers linked to *Salmonella enterica* serovar enteritidis vaccine response in young F1 broiler-cross chicks. *Poult. Sci.* 81, 193–201. doi: 10.1093/ps/81.2.193
- Kaiser, M. G., and Lamont, S. J. (2002). Microsatellites linked to *Salmonella enterica* Serovar Enteritidis burden in spleen and cecal content of young F1 broiler-cross chicks. *Poult. Sci.* 81, 657–663. doi: 10.1093/ps/81.5.657
- Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703
- Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59. doi: 10.1186/1471-2164-14-59
- Kuijl, C., Savage, N. D., Marsman, M., Tuin, A. W., Janssen, L., Egan, D. A., et al. (2007). Intracellular bacterial growth is controlled by a kinase network around PKB/AKT1. *Nature* 450, 725–730. doi: 10.1038/nature06345
- Lambert, W. V., and Knox, C. W. (1932). Selection for resistance to fowl typhoid in the chicken with reference to its inheritance. *Res. Bull.* 12:1.
- Langridge, G. C., Fookes, M., Connor, T. R., Feltwell, T., Feasey, N., Parsons, B. N., et al. (2015). Patterns of genome evolution that have accompanied host adaptation in *Salmonella*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 863–868. doi: 10.1073/pnas.1416707112
- Legarra, A., Calenge, F., Mariani, P., Velge, P., and Beaumont, C. (2011). Use of a reduced set of single nucleotide polymorphisms for genetic evaluation of resistance to *Salmonella* carrier state in laying hens. *Poult. Sci.* 90, 731–736. doi: 10.3382/ps.2010-01260
- Malek, M., Hasenstein, J. R., and Lamont, S. J. (2004). Analysis of chicken TLR4, CD28, MIF, MD-2, and LITAF genes in a *Salmonella* enteritidis resource population. *Poult. Sci.* 83, 544–549. doi: 10.1093/ps/83.4.544
- Mariani, P., Barrow, P. A., Cheng, H. H., Groenen, M. M., Negrini, R., and Bumstead, N. (2001). Localization to chicken chromosome 5 of a novel locus determining salmonellosis resistance. *Immunogenetics* 53, 786–791. doi: 10.1007/s00251-001-0387-7
- Pal, S., Dey, S., Batabyal, K., Banerjee, A., Joardar, S. N., Samanta, I., et al. (2017). Characterization of *Salmonella* Gallinarum isolates from backyard poultry by polymerase chain reaction detection of invasion (invA) and *Salmonella* plasmid virulence (spvC) genes. *Vet. World* 10, 814–817. doi: 10.14202/vetworld.2017.814-817
- Parmar, D., and Davies, R. (2007). Fowl typhoid in a small backyard laying flock. *Vet. Rec.* 160:348. doi: 10.1136/vr.160.10.348

- Parsons, B. N., Humphrey, S., Salisbury, A. M., Mikoleit, J., Hinton, J. C., Gordon, M. A., et al. (2013). Invasive non-typhoidal *Salmonella* typhimurium ST313 are not host-restricted and have an invasive phenotype in experimentally infected chickens. *PLoS Negl. Trop. Dis.* 7:e2487. doi: 10.1371/journal.pntd.0002487
- Poultry Health Scheme Handbook (2013). *Poultry Health Scheme (PHS) Handbook*. Available at: www.defra.gov.uk
- Prasad, K. V., Ao, Z., Yoon, Y., Wu, M. X., Rizk, M., Jacquot, S., et al. (1997). CD27, a member of the tumor necrosis factor receptor family, induces apoptosis and binds to Siva, a proapoptotic protein. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6346–6351. doi: 10.1073/pnas.94.12.6346
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Py, B., Slomianny, C., Auberger, P., Petit, P. X., and Benichou, S. (2004). Siva-1 and an alternative splice form lacking the death domain, Siva-2, similarly induce apoptosis in T lymphocytes via a caspase-dependent mitochondrial pathway. *J. Immunol.* 172, 4008–4017. doi: 10.4049/jimmunol.172.7.4008
- Redmond, S. B., Chuammitri, P., Andreasen, C. B., Palic, D., and Lamont, S. J. (2011). Genetic control of chicken heterophil function in advanced intercross lines: associations with novel and with known *Salmonella* resistance loci and a likely mechanism for cell death in extracellular trap production. *Immunogenetics* 63, 449–458. doi: 10.1007/s00251-011-0523-y
- Revolledo, L. (2018). Vaccines and vaccination against fowl typhoid and pullorum disease: an overview and approaches in developing countries. *J. Appl. Poult. Res.* 27, 279–291. doi: 10.3382/japr/pfx066
- Shivaprasad, H. L. (2000). Fowl typhoid and pullorum disease. *Rev. Sci. Tech.* 19, 405–424. doi: 10.20506/rst.19.2.1222
- Steele-Mortimer, O., Knodler, L. A., Marcus, S. L., Scheid, M. P., Goh, B., Pfeifer, C. G., et al. (2000). Activation of Akt/protein kinase B in epithelial cells by the *Salmonella* typhimurium effector sigD. *J. Biol. Chem.* 275, 37718–37724. doi: 10.1074/jbc.M008187200
- Taniguchi, T., Harada, H., and Lamphier, M. (1995). Regulation of the interferon system and cell growth by the IRF transcription factors. *J. Cancer Res. Clin. Oncol.* 121, 516–520. doi: 10.1007/BF01197763
- Thanh-Son, T., Catherine, B., Nigel, S., Mark, F., Pete, K., Elisabeth Le, B. D., et al. (2012). A maximum likelihood QTL analysis reveals common genome regions controlling resistance to *Salmonella* colonization and carrier-state. *BMC Genomics* 13:198. doi: 10.1186/1471-2164-13-198
- Tilquin, P., Barrow, P. A., Marly, J., Pitel, F., Plisson-Petit, F., Velge, P., et al. (2005). A genome scan for quantitative trait loci affecting the *Salmonella* carrier-state in the chicken. *Genet. Sel. Evol.* 37, 539–561. doi: 10.1186/1297-9686-37-6-539
- Weerasooriya, K., Fernando, P. S., Liyanagunawardena, N., Wijewardena, G., Wijemuni, M. I., and Samarakoon, S. (2017). Natural resistance of Sri Lankan village chicken to *Salmonella* gallinarum infection. *Br. Poult. Sci.* 58, 644–648. doi: 10.1080/00071668.2017.1376034
- Wigley, P. (2017). *Salmonella enterica* serovar Gallinarum: addressing fundamental questions in bacteriology sixty years on from the 9R vaccine. *Avian Pathol.* 46, 119–124. doi: 10.1080/03079457.2016.1240866
- Wigley, P., Hulme, S. D., Bumstead, N., and Barrow, P. A. (2002). In vivo and in vitro studies of genetic resistance to systemic salmonellosis in the chicken encoded by the SAL1 locus. *Microbes Infect.* 4, 1111–1120. doi: 10.1016/S1286-4579(02)01635-0
- Yamaoka, K., Saharinen, P., Pesu, M., Holt, V. E. III, Silvennoinen, O., and O'shea, J. J. (2004). The Janus kinases (Jaks). *Genome Biol.* 5:253. doi: 10.1186/gb-2004-5-12-253
- Yoshida, I., and Azuma, M. (1992). Function, molecular structure and gene expression of interferons. *Nihon Rinsho* 50, 1845–1853.
- Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* 11, 407–409. doi: 10.1038/nmeth.2848

Conflict of Interest Statement: JF was employed by Hy-Line International.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Psifidi, Russell, Matika, Sánchez-Molano, Wigley, Fulton, Stevens and Fife. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Sequence Characterization of *DSG3* Gene to Know Its Role in High-Altitude Hypoxia Adaptation in the Chinese Cashmere Goat

Chandar Kumar^{1,2}, Shen Song¹, Lin Jiang¹, Xiaohong He¹, Qianjun Zhao¹, Yabin Pu¹, Kanwar Kumar Malhi³, Asghar Ali Kamboh³ and Yuehui Ma^{1*}

¹ The Key Laboratory for Farm Animal Genetic Resources and Utilization of Ministry of Agriculture of China, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, ² Department of Animal Breeding and Genetics, Faculty of Animal Husbandry and Veterinary Sciences, Sindh Agriculture University, Tando Jam, Pakistan, ³ Department of Veterinary Microbiology, Faculty of Animal Husbandry and Veterinary Science, Sindh Agriculture University, Tando Jam, Pakistan

OPEN ACCESS

Edited by:

John Anthony Hammond,
Pirbright Institute (BBSRC),
United Kingdom

Reviewed by:

Shaojun Liu,
Hunan Normal University, China
Keith Ballingall,
Moredun Research Institute,
United Kingdom

*Correspondence:

Yuehui Ma
yuehui.ma@263.net

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 05 June 2018

Accepted: 29 October 2018

Published: 19 November 2018

Citation:

Kumar C, Song S, Jiang L, He X, Zhao Q, Pu Y, Malhi KK, Kamboh AA and Ma Y (2018) Sequence Characterization of *DSG3* Gene to Know Its Role in High-Altitude Hypoxia Adaptation in the Chinese Cashmere Goat. *Front. Genet.* 9:553. doi: 10.3389/fgene.2018.00553

The Tibetan cashmere goat is one of the main goat breeds used by people living in the plateau. It exhibits the distinct phenotypic characteristics observed in lowland goats, allowing them to adapt to the challenging conditions at high altitudes. It provides an ideal model for understanding the genetic mechanisms underlying high-altitude adaptation and hypoxia-related diseases. Our previous exome sequencing of five Chinese cashmere breeds revealed a candidate gene, *DSG3* (Desmoglein 3), responsible for the high-altitude adaptation of the Tibetan goat. However, the whole *DSG3* gene (44 kbp) consisting of 16 exons in the goat genome was not entirely covered by the exome sequencing. In this study, we resequenced all the 16 exons of the *DSG3* gene in ten Chinese native goat populations. Twenty-seven SNP variants were found between the lowland and highland goat populations. The genetic distance (F_{ST}) of significant SNPs between the lowland and highland populations ranged from 0.42 to 0.58. By using correlation coefficient analysis, linkage disequilibrium, and haplotype network construction, we found three non-synonymous SNPs (R597E, T595I, and G572S) in exon 5 and two synonymous SNPs in exons 8 and 16 in *DSG3*. These mutations significantly segregated high- and low-altitude goats in two clusters, indicating the contribution of *DSG3* to the high-altitude hypoxia adaptation in the Tibetan goat.

Keywords: *DSG3*, exons, SNPs, Tibetan goat, hypoxia, high-altitude adaptation

INTRODUCTION

Low oxygen or hypoxia is the hardest environmental challenge for humans and animals existing at high altitude. Whole-genome resequencing analyses have been carried out to investigate the genetic basis of the hypoxia adaptation in many domestic animals, such as yak (Qiu et al., 2012), Tibetan pig (Ai et al., 2014), dog (Li et al., 2014), and birds (Cai et al., 2013). Various candidate genes have been identified for high-altitude adaptation, including *EPAS1* (endothelial PAS domain protein 1) and *HBB* (hemoglobin beta) (Gou et al., 2014; Fan et al., 2015; Song et al., 2016). In prior studies, both *EPAS1* and *HBB* revealed six non-synonymous mutations potentially affecting the gene function and influencing high-altitude hypoxic adaptation in dogs (Fan et al., 2015). These studies proved that the domestic animals are a useful animal model to explore high-altitude adaptation.

The domestic goat is found at sea level (30 m) to the high plateau (4700 m). In contrast to its low-altitude counterpart, the Tibetan goat has unique anatomical and physiological characteristics such as higher hemoglobin concentration, a larger heart, and bigger lungs that equip it to live at high altitudes (Li et al., 2004b). Our previous exome-sequencing analysis also reveals that the cardio-vascular system-related genes may play a crucial role in high-altitude adaptation (Song et al., 2016). Among the candidate genes under strong selection, the *DSG3* (Desmoglein 3) loci was the most significant. *DSG3* is a desmoglein gene located in a cluster on goat chromosome 24 containing 44 kbp that encodes 16 exons. It is expressed in stratifying epithelia, but its precise role is not fully understood in the cardiovascular system (Delva et al., 2009). Here, we extended our previous analyses on large goat populations from different altitudes to identify if the *DSG3* candidate gene has a role in high-altitude hypoxia adaptation. By using correlation coefficient analysis, linkage disequilibrium, and haplotype network construction, we explored both the non-synonymous and synonymous mutation sites of the *DSG3* gene. These mutations significantly segregated goat populations between high- and low-altitude groups.

MATERIALS AND METHODS

Populations Sampling

A total of 125 Chinese cashmere goats were used for Sanger sequencing and covered the entire genomic region of *DSG3* (44 kbp) consisting of 16 exons. Isolated genomic DNA used from Ritu (RT, 4700 m), Bange (BG, 400 m), Nanjiang (NJ, 1700 m), Inner Magnolia (IM, 1500 m), and Liaoning (LN, 30 m) cashmere goats selected from 330 individuals was analyzed previously for Exome Sequencing by the Illumina HiSeq2000 Platform (Song et al., 2016), which were randomly selected from 4 different locations (Table 1 and Supplementary Table S1 and Supplementary Figure S1). Furthermore, five goat populations including Dulan (DL, 3000 m), Hanshan (HS, 1000 m), Guangfeng (GF, 500 m), Hainan (HN, 120 m), and Changjiangsanjiaozhou (CZ, 50 m) were selected from five different geographical locations in China (Table 1 and Supplementary Table S1 and Supplementary Figure S1). The tissue samples of these goat populations were taken by cutting ear tissue, using an ear-cutting instrument without applying an anesthetic or analgesic agent. After the collection, the tissue samples were preserved in a tube containing 75% alcohol, which was kept in the ice box during transportation to the laboratory. The tissue samples were collected from all the animals undergoing the experiment according to the recommendations and guideline of the Institute of Animal Sciences (IAS, CAAS), with full approval from the Animal Care and Use Committee of Chinese Academy of Agricultural Sciences and the Ministry of Agriculture of the People's Republic of China.

Genomic DNA was isolated from ear tissue using a modified commercial Kit protocol (Promega, the Wizard®). The quality and integrity of the purified DNA from each goat per sample location were determined using a Nanodrop 2000 (Thermo

Fisher Scientific, DE) and pooled according to the concentration bases. The validations in the larger goat population were carried out by following sample data and locations. Highland group consisted of Ritu (RT, $n = 17$) and Bange (BG, $n = 17$) goats from the Ritu and Bange region of Tibet. The lowland group consisted of Dulan (DL, $n = 8$) and Nanjiang (NJ, $n = 14$) cashmere goats from Qinhai, Dulan, and Aksu areas of Xinjiang region, Inner Magnolia (IM, $n = 13$) from the Earlangshan region of Inner Mongolia, Hanshan (HS, $n = 8$) from Inner Magnolia city, Guangfeng (GF, $n = 15$) from Shangrao, Jiangxi, Hainan (HN, $n = 8$) from Hainan city, Changjiangsanjiaozhou (CZ, $n = 13$) from Changjiangsanjiaozhou city and Liaoning (LN, $n = 15$) cashmere goats from Gai Zhou, Liaoning, as shown in Table 1, Supplementary Table S1 and Supplementary Figure S1.

Polymerase Chain Reaction (PCR) and Single-Nucleotide Polymorphisms (SNPs) detection

Sixteen pairs of primers were used for *DSG3* gene amplification and their sequences are shown in Table 2. Polymerase chain reaction (PCR) was optimized with a 50 ng DNA template using 2xTaq master mix (BioMed: Beijing Co., Ltd.), supplemented with ddH2O solutions and 10 pmol/l of forward and reverse primers and in a 30 μ l mixture. The PCR was performed by initial denaturation for 5 min at 95°C, followed by 40 cycles at 95°C for 30 s, annealing at 60–63°C for 30 s and an extension at 72°C for 45 s, and a final extension at 72°C for 5 min and stored at 4°C. The PCR products were analyzed by running on 1.5% of gel electrophoresis that was made by a mixture of 0.75 g Regular Agarose Biowest (Gene Company) and 50 ml of 1x TAE buffer. The mixture was also added with 5 μ l gold-view chemical for viewing the gel on UV transilluminator. The electrophoresis tank (Bio-rad Sub-Cell® GT) and the PowerPac Basic (Bio-rad®) were used for running the electrophoresis. Five microliters of the PCR products were mixed with 10X loading dye, which is composed of 0.25% Bromophenol blue, 0.25% Xylene Cyanol FF, 15% Ficoll, and H2O. The mixture was loaded into each running well of 1.5% gel. The electrophoresis was run under 1X TAE buffer, carried out at a 75W constant, 147 V for 40 min. When electrophoresis was finalized, the gels were set on the UV transilluminator. All the purified PCR products were directly sequenced by the Sanger Sequence methods, using the ABI 3730 sequence analyzer (Applied Bio-System, United States) in both the directions, and using the forward primer. The sequences were assembled and analyzed for single-nucleotide polymorphism (SNPs), using the Seq Man program of Laser gene software. The SNPs were detected from Primer 4, Primer 5, Primer 6, Primer 8, Primer 9, Primer 11, Primer 14, and Primer 16 in the *DSG3* gene through DNA pooled strategies. Furthermore, these primers were used in large populations on an individual basis to identify polymorphism and the validation of the SNPs.

Statistical Analyses

Pairwise genetic distance (F_{ST}) was calculated between High-Lowland and High-Midland and Mid-lowland goat

TABLE 1 | Sample locations and major allele frequency of three non-synonymous variants and two synonymous mutations in *DSG3* gene.

| Sample group | Sample location | Altitude (m) | SIZE | SNP1 (R597E) | SNP2 (T595I) | SNP3 (G572S) | SNP4 | SNP5 |
|--------------|-----------------------------|--------------|-------|--------------|--------------|--------------|------|------|
| LN | Gai Zhou, Liaoning | 30 m | 15.00 | 0.37 | 0.37 | 0.00 | 0.23 | 0.20 |
| CZ | Changjiangsanjiaozhou | 50 m | 13.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| HN | Hainan | 120 m | 8.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 |
| GF | Shangrao, Jiangxi | 500 m | 15.00 | 0.06 | 0.06 | 0.00 | 0.00 | 0.00 |
| HS | Inner Mongolia | 1000 m | 8.00 | 0.00 | 0.00 | 0.69 | 0.00 | 0.00 |
| IM | Earlangshan, Inner Mongolia | 1500 m | 13.00 | 0.65 | 0.73 | 0.23 | 0.58 | 0.62 |
| NJ | Aksu, Xinjiang | 1700 m | 14.00 | 0.23 | 0.23 | 0.69 | 0.19 | 0.18 |
| DL | Qinhai, haixizhou, Dula | 3000 m | 8.00 | 0.56 | 0.56 | 0.69 | 0.56 | 0.56 |
| BG | Bange, Tibet | 4000 m | 17.00 | 0.67 | 0.67 | 0.67 | 0.67 | 0.69 |
| RT | Ritu, Tibet | 4700 m | 17.00 | 0.97 | 0.97 | 0.97 | 0.97 | 1.00 |

TABLE 2 | Sequences of the primers.

| PCR primer | 5' → 3' | PCR region | Anneal temp | Product length |
|----------------------|--------------------------|------------|-------------|----------------|
| Exon-1F <i>DSG3</i> | GGCCTCTTTTGCTTGTGAGAAA | 25786301 | 61 | 1571 |
| Exon-1R <i>DSG3</i> | CAGTCACGATCAAACGTTGAGTT | 25784966 | | |
| Exon-2F <i>DSG3</i> | CAGTAGCCGGATTTTGGCAG | 25788145 | 60 | 528 |
| Exon-2R <i>DSG3</i> | ACTTGGCATGTGCTCTTGGA | 25787897 | | |
| Exon-3F <i>DSG3</i> | ACCGATTCATTGAAACCTGT | 25791050 | 61 | 412 |
| Exon-3R <i>DSG3</i> | CTCCTGACCATCAAGGACCA | 25790985 | | |
| Exon-4F <i>DSG3</i> | GGTCCTTGATGGTCAGGAGTC | 25791491 | 64 | 331 |
| Exon-4R <i>DSG3</i> | TCCAACAATCGCCCTGTCAA | 25791351 | | |
| Exon-5F <i>DSG3</i> | AACATACACGACCTGCTCTGC | 25794845 | 63 | 600 |
| Exon-5R <i>DSG3</i> | AACCCCAACAGCCCTCATAA | 25794593 | | |
| Exon-6F <i>DSG3</i> | TGCTCCCAATGGACTCCC | 25796339 | 63 | 365 |
| Exon-6R <i>DSG3</i> | GCACAACCTGCTCTGGATAGTTC | 25796111 | | |
| Exon-7F <i>DSG3</i> | GCAGTGATAAAGGTAATGTGTTAT | 25798279 | 60 | 422 |
| Exon-7R <i>DSG3</i> | GCAGGAAATTTAAACCCAGGCT | 25798137 | | |
| Exon-8F <i>DSG3</i> | ATGAGGCAACTTCGCCAGTG | 25799330 | 60 | 337 |
| Exon-8R <i>DSG3</i> | GAGTTCTCTGGTCTCTTACCTG | 25799057 | | |
| Exon-9F <i>DSG3</i> | TGTGCTTAAACAGGCACACT | 25802312 | 59 | 599 |
| Exon-9R <i>DSG3</i> | AGGGACCAGCAAAGAGACAAT | 25802125 | | |
| Exon-10F <i>DSG3</i> | GCAAGCCAGCAAATCCTGAA | 25802738 | 60 | 415 |
| Exon-10R <i>DSG3</i> | CCCCACAGGCTTTGTCTACG | 25802604 | | |
| Exon-11F <i>DSG3</i> | ACCCCTCTCTCCTCCCTCTTT | 25803899 | 59 | 411 |
| Exon-11R <i>DSG3</i> | GTGACACGTAGATTTTGGCAC | 25803733 | | |
| Exon-12F <i>DSG3</i> | TGGCAGTGGGTAAGTAGCTC | 25804674 | 60 | 514 |
| Exon-12R <i>DSG3</i> | GAAGGGGCAACTTGACAAAC | 25804524 | | |
| Exon-13F <i>DSG3</i> | ATTCACGCACTGATACCTTCT | 25805338 | 59 | 324 |
| Exon-13R <i>DSG3</i> | GACTTGAGAGTGTGGAGCAC | 25805217 | | |
| Exon-14F <i>DSG3</i> | TGTTCTATGGACCCCTCAATCT | 25807166 | 59 | 400 |
| Exon-14R <i>DSG3</i> | TGCATGATAGTGTGAGGAGAG | 25807028 | | |
| Exon-15F <i>DSG3</i> | CACAAATACCGTTTATCTGTTGA | 25807700 | 51 | 366 |
| Exon-15R <i>DSG3</i> | TGAACAGCAGAATGTAAATCAAA | 25807664 | | |
| Exon-16F <i>DSG3</i> | ATCTGGCTTGGTTTCCATGTG | 25817374 | 60 | 472 |
| Exon-16R <i>DSG3</i> | GAGCAAAAGTGATGCAAGCGT | 25817194 | | |

populations to find out the genetic divergence as described by Weir and Cockerham (1984). Fisher's exact test was estimated respectively, to observe a statistical significance. LD-block structure and parameter (D' and r^2) were figured using Haploview 4.2 software (Barrett et al., 2005). To explore the signature of selection and evolutionary change, the six SNPs

(four Non-Synonymous and two Synonymous variants) were analyzed for the diversity of haplotype and their frequency was estimated using phase 2.1 (Stephens et al., 2001; Crawford et al., 2004). To obtain a reliable result, we used option – X 10 to increase final runs time and – C option for low- and high-land goat groups to ensure that any similarity to

the haplotype is not taken into account. The haplotype was embedded in DNASP Version5 (Librado and Rozas, 2009) for network construction and phylogenetic diversity among the identified haplotypes was inferred through a median-joining network analysis using the method recommended by Bandelt et al. (1999). To detect the functions of three major non-synonymous SNPs in the *DSG3* gene, the SIFT software¹ has been used. The protein sequence of Tibetan goats was compared with the reference protein sequence of goats desmoglein-3. The cutoff value in the SIFT program is a tolerance score of ≥ 0.05 . The higher the tolerance score, the lower the expected functional impact of a particular amino acid substitution.

RESULTS

Allele Frequency and Genetic Differentiation

In this study, we sampled 128 Chinese indigenous goats belonging to ten populations living across a wide distribution of altitudes. The ten populations ranged from sea level (30 m) to high altitude (4700 m) (Table 1 and Supplementary Table S1). To find out the polymorphism and to measure allele frequencies of the *DSG3* gene accurately, we resequenced the entire genomic region of the *DSG3* gene. Among the 27 SNPs that were identified in *DSG3*, five SNPs showed the most significant differentiation in the analysis of allele frequency and global F_{ST} , including three non-synonymous SNPs, SNP1 (R597E, Chr24: 25794694, exon 5), SNP2 (T595I, Chr24: 25794695, exon 5), and SNP3 (G572S, Chr24: 25794771, exon 5), and two synonymous SNPs, SNP4 (Chr24: 25799255, exon 8), and SNP5 (Chr24: 25817330, exon 16) (Table 3). A striking change in allele frequency with elevation in altitude was observed in goat population. For example, low R597E allele frequencies in *DSG3* were found in lowland populations including CZ (0.00, 50m), HN (0.00, 120 m), and HS (0.00, 1000 m), but were elevated in the IM (0.65, 1500 m), DL (0.56, 3000 m), BG (0.67, 4000 m), and reached the highest frequency in the RT (0.97, 4700 m) population (Table 1, Supplementary Table S2, and Figure 1A). The other four SNPs in *DSG3* exhibited a similar pattern of change (Table 1, Supplementary Table S2, and Figures 1, 2). All these results indicated that the frequencies of the five variants in the *DSG3* locus remain rare in the

lowland goat population but increased with the elevation of altitude.

Using allele frequencies of the three non-synonymous and two synonymous SNPs measured above, we measured the correlation co-efficiency of variant allele frequency per population versus the altitude of sampling location using Graph Pad Prism 5 software. The result of this analysis showed a significant positive linear correlation between the elevated altitude and the frequency of a variant allele in *DSG3* (Table 3 and Supplementary Table S3). The linear correlation coefficient between the elevated altitude and the frequencies of variant allele was $r = 0.85$, $P < 0.05$ for SNP1 (Figure 1B), $r = 0.83$, $P < 0.05$ for SNP2 (Figure 1D), $r = 0.84$, $P < 0.05$ for SNP3 (Figure 1F). The linear correlation coefficient was $r = 0.91$, $P < 0.05$ for SNP 4 (Figure 2B), and $r = 0.90$, $P < 0.05$ for SNP5 (Table 3, Supplementary Table S3, and Figure 2D). The significant positive correlation between the elevated altitudes and the variant allele frequency at the *DSG3* locus in ten Chinese native goat populations strongly suggested that the candidate gene *DSG3* potentially contributed to the high-altitude adaptation in goats.

To further understand the genetic differentiation that occurred between the lowland and highland goat populations, we first estimated the global F_{ST} value for all the identified loci. The mean F_{ST} value was 0.24, indicating the high genetic divergence in these ten Chinese native goat populations, which is mainly attributed to the divergence between the lowland and highland goat populations. The global F_{ST} of all the 27 loci ranged from 0.0 to 0.61 (Supplementary Table S2). The three most divergent non-synonymous SNPs had the value that ranged from 0.51 (SNP1) and 0.52 (SNP2) to 0.55 (SNP3), whereas the top two synonymous SNPs had a global F_{ST} of 0.47 (SNP4) and 0.61 (SNP5) (Table 3 and Supplementary Table S2). Subsequently, the pairwise genetic distance (F_{ST}) was measured between High-Low, High-Mid, and Mid-Low altitude goat populations (Table 4). The F_{ST} values of non-synonymous variants between high-altitude and low-altitude populations were 0.42 (SNP2, $P < 0.001$) and 0.45 (SNP1/SNP3, $P < 0.001$), which reflected a remarkable genetic differentiation (Table 4). Similarly, the F_{ST} values of the two synonymous substitutes were 0.47 (SNP4, $P < 0.001$) and 0.58 (SNP5, $P < 0.001$). These results showed high genetic differentiation between High-Low altitude goat populations compared with the High-Mid and Mid-Low altitude goat populations (Table 4). Taken together, both of global F_{ST} values and pairwise genetic distance (F_{ST}) values demonstrate

¹<http://sift.bii.a-star.edu.sg/>

TABLE 3 | Correlation and global F_{ST} value of five mutation loci in *DSG3* gene.

| SNPs | Genomic location | Reference allele | Mutation allele | Pearson r | G. F_{ST} values |
|--------------|------------------|------------------|-----------------|-----------|--------------------|
| SNP1 (R597E) | Chr24: 25794694 | C | T | 0.8529 | 0.50 |
| SNP2 (T595I) | Chr24: 25794695 | G | C | 0.8333 | 0.55 |
| SNP3 (G572S) | Chr24: 25794771 | T | C | 0.8415 | 0.50 |
| SNP4 | Chr24: 25799255 | G | A | 0.9061 | 0.47 |
| SNP5 | Chr24: 25817330 | G | A | 0.9000 | 0.61 |

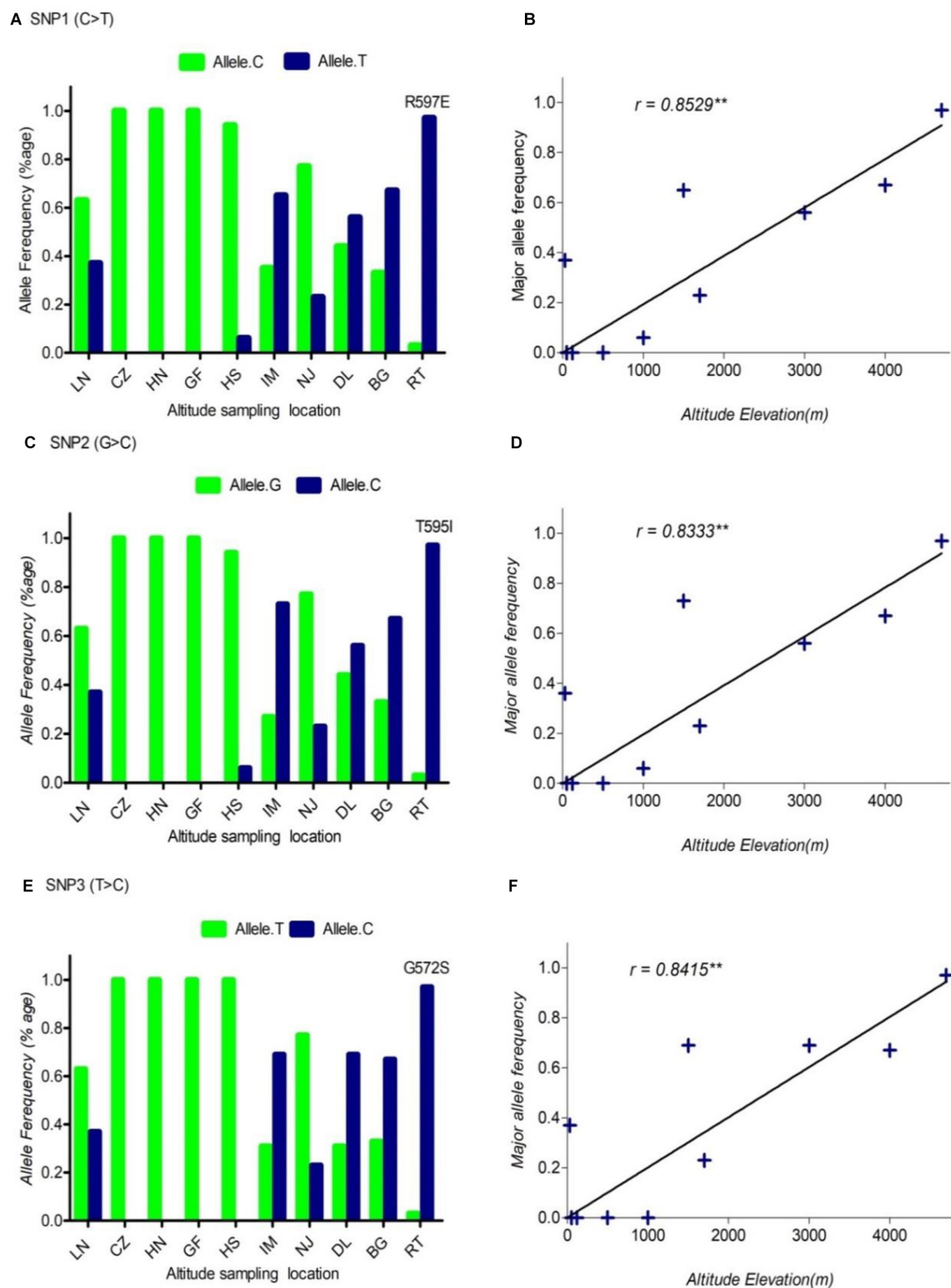
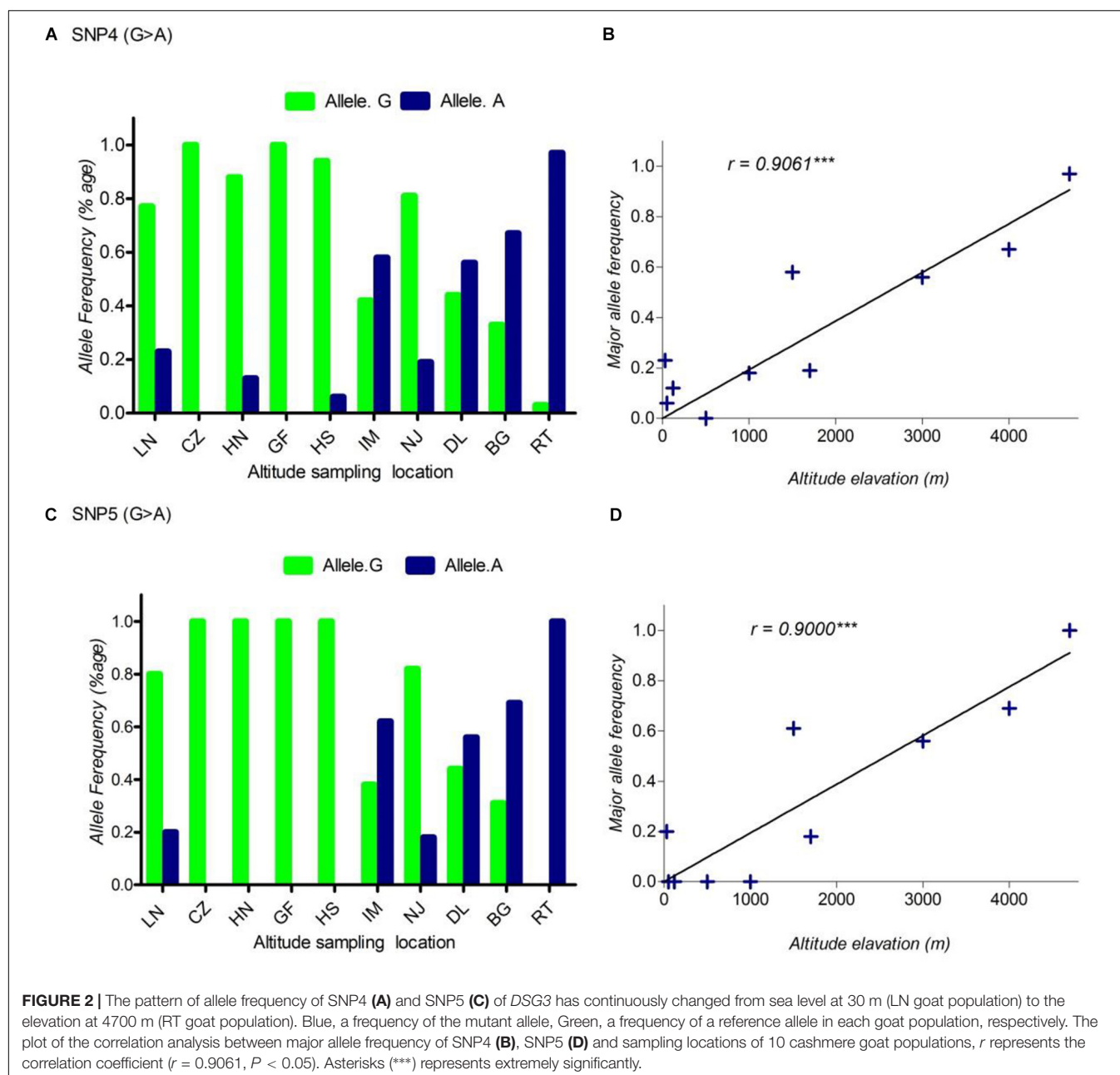


FIGURE 1 | The pattern of allele frequency of SNP1 (A), SNP2 (C) and SNP3 (E) of *DSG3* has continuously changed from sea level at 30 m (LN goat population) to the elevation at 4700 m (RT goat population). Blue, a frequency of the mutant allele, Green, a frequency of a reference allele in each goat population, respectively. The plot of the correlation analysis between major allele frequency of SNP1 (B), SNP2 (D) and SNP3 (F) and sampling locations of 10 cashmere goat populations, r represents the correlation coefficient ($r = 0.8529$, $P < 0.05$). Asterisks (**) represents very significantly.



a remarkable population differentiation between the high- and low-altitude goat populations.

Linkage Disequilibrium (LD) and Haplotype Network Analysis

The LD haplotype pattern was used to analyze the *DSG3* loci (Supplementary Figure S2). We constructed LD blocks using the five significant SNPs and calculated D' , r^2 , and an algorithm of the odds (LOD). The variant allele of these five SNPs was tightly linked as a Haplo-Block on chromosome 24 in highland goat populations (RT-BG), with the three non-synonymous SNPs in the main position and showed strong

correlation by $D' = 1.0 - 1.0$, $r^2 = 0.63 - 1.00$, and $\text{LOD} = 5.19 - 9.82$ (Table 5 and Figure 3A). In contrast, the linkage of three non-synonymous SNPs and two synonymous SNPs was loosely found in lowland goat populations (DL-NJ-AB-LN-GF-CZ-HN-HS). However, the LD level of SNPs varied from $D' = 0.58 - 1$, $r^2 = 0.10 - 0.94$, and $\text{LOD} = 2.85 - 31.71$ (Table 5 and Figure 3B). The haplotype pattern was analyzed to detect the effect of natural selection on selected genes, including *DSG3*. Twelve haplotypes were obtained from four non-synonymous (SNP1/R392Q, SNP2/T595I, SNP3/G572S, and T480N) and two from synonymous SNPs (SNP4 and SNP5) from 128 goats from lowland and highland goat populations. The top three haplotypes were detected: HL1, HL2, and L7 with a

TABLE 4 | Pairwise genetic distances (F_{ST}) & Fisher exact test (p -value) calculated between highland and reference goat populations.

| Chromosome position | SNPs | M allele | Anno. | H_M (F_{ST}) | P-value | H_L (F_{ST}) | P-value | M_L (F_{ST}) | P-value |
|---------------------|-------|----------|-------|---------------------|----------|---------------------|----------|---------------------|----------|
| Chr24:25791412 | | T | | 0.06 | 0.00 | 0.08 | 1.13E-04 | 0.00 | 0.88 |
| Chr24:25794517 | | G | | 0.05 | 0.00 | 0.12 | 1.30E-07 | 0.02 | 0.03 |
| Chr24:25794535 | | A | | 0.05 | 0.00 | 0.01 | 0.12 | 0.03 | 0.13 |
| Chr24:25794564 | | A | | 0.03 | 0.06 | 0.00 | 1.00 | 0.03 | 0.06 |
| Chr24:25794694 | SNP1* | T | R597E | 0.25 | 3.03E-12 | 0.45 | 1.22E-20 | 0.04 | 0.02 |
| Chr24:25794695 | SNP2* | C | T595I | 0.25 | 3.03E-12 | 0.42 | 2.17E-19 | 0.03 | 0.04 |
| Chr24:25794700 | | A | | 0.03 | 0.06 | 0.00 | 0.25 | 0.03 | 0.68 |
| Chr24:25794771 | SNP3* | C | G572S | 0.20 | 3.24E-10 | 0.45 | 5.25E-20 | 0.06 | 0.00 |
| Chr24:25794881 | | A | | 0.10 | 1.40E-05 | 0.18 | 3.53E-09 | 0.01 | 0.13 |
| Chr24:25794882 | | A | | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 1.00 |
| Chr24:25794934 | | A | | 0.00 | 1.00 | 0.12 | 1.30E-06 | 0.11 | 1.30E-06 |
| Chr24:25796278 | | C | | 0.10 | 0.00 | 0.19 | 2.75E-09 | 0.02 | 0.04 |
| Chr24:25796283 | | C | | 0.04 | 0.02 | 0.06 | 0.00 | 0.00 | 0.41 |
| Chr24:25799144 | | G | | 0.03 | 0.06 | 0.00 | 1.00 | 0.03 | 0.21 |
| Chr24:25799255 | SNP4* | G | | 0.25 | 1.73E-13 | 0.47 | 2.08E-22 | 0.05 | 0.01 |
| Chr24:25799281 | | G | | 0.01 | 1.00 | 0.03 | 0.01 | 0.06 | 0.01 |
| Chr24:25801885 | | C | | 0.01 | 0.06 | 0.11 | 5.71E-08 | 0.07 | 0.00 |
| Chr24:25802080 | | T | | 0.01 | 1.00 | 0.03 | 0.03 | 0.06 | 0.03 |
| Chr24:25802106 | | G | | 0.00 | 1.00 | 0.01 | 0.50 | 0.01 | 0.50 |
| Chr24:25803941 | | C | | 0.08 | 0.01 | 0.27 | 1.01E-10 | 0.09 | 0.00 |
| Chr24:25803953 | | G | | 0.01 | 0.50 | 0.00 | 1.00 | 0.00 | 1.00 |
| Chr24:25807002 | | T | | 0.00 | 2.19E-13 | 0.00 | 7.22E-17 | 0.00 | 0.35 |
| Chr24:25807209 | | T | | 0.06 | 0.01 | 0.14 | 6.14E-07 | 0.02 | 0.02 |
| Chr24:25807240 | | T | | 0.00 | 1.00 | 0.09 | 7.26E-06 | 0.09 | 7.27E-06 |
| Chr24:25817330 | SNP5* | A | | 0.31 | 8.94E-16 | 0.58 | 4.25E-26 | 0.07 | 0.01 |
| Chr24:25817361 | | G | | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.50 |
| Chr24:25817366 | | G | | 0.03 | 0.06 | 0.02 | 0.10 | 0.00 | 0.80 |

Top five SNPs are shown with asterisk.

TABLE 5 | Pairwise linkage disequilibrium of the non-synonymous and synonymous SNPs of *DSG3* in the highland and lowland goat populations.

| SNPS | | Highland goat | Lowland goat | Highland goat | Lowland goat | Highland goat | Lowland goat |
|--------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|
| SNP1 | SNP2 | D' | | r^2 | | LOD | |
| SNP1 (C > T) | SNP2 (G > C) | 1.00 | 1.00 | 1.00 | 0.94 | 9.82 | 31.71 |
| SNP1 (C > T) | SNP3 (T > C) | 1.00 | 1.00 | 1.00 | 0.91 | 9.82 | 30.46 |
| SNP1 (C > T) | SNP4 (G > A) | 1.00 | 0.93 | 1.00 | 0.79 | 9.82 | 22.95 |
| SNP1 (C > T) | SNP5 (G > A) | 1.00 | 0.96 | 0.63 | 0.73 | 5.19 | 21.95 |
| SNP2 (G > C) | SNP3 (T > C) | 1.00 | 0.94 | 1.00 | 0.86 | 9.82 | 27.56 |
| SNP2 (G > C) | SNP4 (G > A) | 1.00 | 0.93 | 1.00 | 0.74 | 9.82 | 21.87 |
| SNP2 (G > C) | SNP5 (G > A) | 1.00 | 1.00 | 0.63 | 0.75 | 5.19 | 24.01 |
| SNP3 (T > C) | SNP4 (G > A) | 1.00 | 0.93 | 1.00 | 0.71 | 9.82 | 21.12 |
| SNP3 (T > C) | SNP5 (G > A) | 1.00 | 0.96 | 0.63 | 0.67 | 5.19 | 20.58 |
| SNP4 (G > A) | SNP5 (G > A) | 1.00 | 0.96 | 0.63 | 0.84 | 5.19 | 25.26 |

frequency of 42%, 40%, and 30%, respectively (**Figure 3C**). Among these haplotypes, HL1 showed a remarkably higher haplotype frequency of a mutant allele (TCCCAA) in highland population than in lowland populations (18.0%). However, HL2 indicated a much lower haplotype frequency of the reference allele (CGTAGG) in the highland (7.0%) than in the

lowland population (42%). L7 showed a haplotype frequency of (CGTCGG) in lowland (31%) and indicated a separate haplotype cluster in lowland populations. These results suggested that the haplotype TCCCAA of the *DSG3* gene has been selected during the high-altitude adaptation of the Tibetan goat.

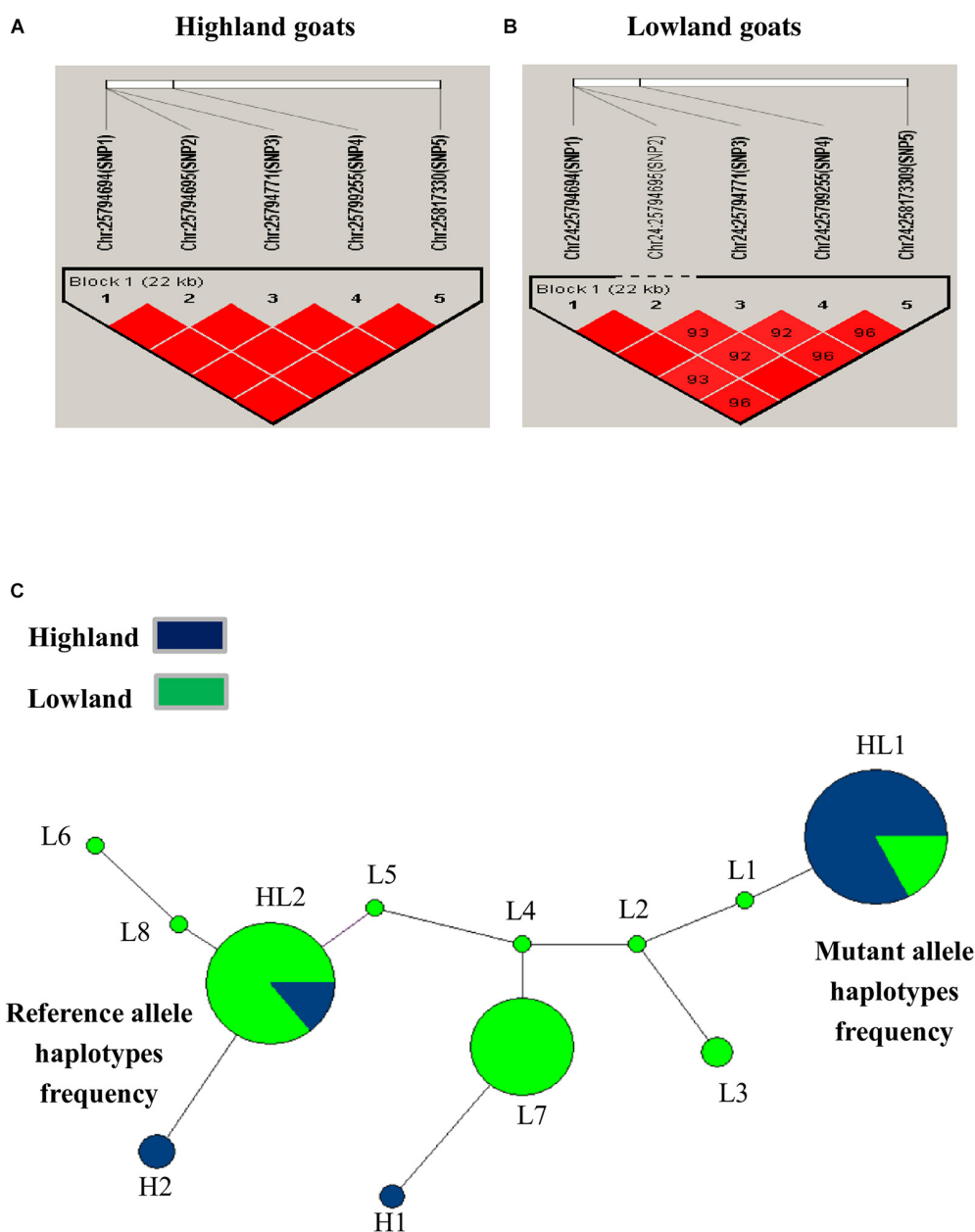


FIGURE 3 | (A) Linkage disequilibrium of non-synonymous variants (SNP1, SNP2, and SNP3) and synonymous substitutions (SNP4 and SNP5) in highland goat populations; RT and BG. **(B)** Linkage disequilibrium of non-synonymous variants (SNP1, SNP2, and SNP3) and synonymous substitutions (SNP4 and SNP5) in lowland goat populations; LN, CZ, HN, GF, HS, IM, NJ, and DL. Strengths of the LD between two SNPs were indicated by the color scheme (a bright red diamond indicates a strong correlation) by the pairwise D' values given as the percentage in each diamond (a bright red diamond indicates $D' = 1.00$). **(C)** Haplotype networks of non-synonymous variants (SNP1, SNP2, and SNP3, T480N) and synonymous substitutions (SNP4 and SNP5) among lowland and highland goat populations. Each node represents a haplotype, with the size of the circle proportional to frequency. The length of a branch is proportional to the number of nucleotide differences. Circles are color code according to the population. Blue: highland (RT and BG), Green: lowland (LN, CZ, HN, GF, HS, IM, NJ, and DL), abbreviation: HL, high and low altitude; H, high altitude; L, low altitude.

Prediction of Three Non-synonymous SNPs

An online tool SIFT was used to investigate the potential impact of the non-synonymous substitutions on protein structure and function. SIFT aligns a query sequence and a subject amino acid

sequence to determine the effect of an amino acid substitution. SIFT software takes protein FASTA sequences as input and is aligned with PSI-BLAST. In our result analysis, we found three nsSNPs to be predicted, as shown in **Table 6**, by using SIFT software.

TABLE 6 | Predications of non-synonymous SNPs function by SIFT software.

| nsSNPs | Positions | Reference allele | Mutant allele | Amino acid changed | Predictions | Score |
|--------|-----------|------------------|---------------|--------------------|-------------------------|-------|
| SNP1 | 597 | C | T | R597E | Affect protein function | 0.000 |
| SNP2 | 591 | G | C | T595I | Affect protein function | 0.000 |
| SNP3 | 572 | T | C | G572S | Affect protein function | 0.000 |

Amino acid substitutions changed in Tibetan goat protein sequences may have been predicted to modify the function of DSG3 gene (low certainty). The cutoff value in the SIFT program is a tolerance score of ≥ 0.05 . The higher the tolerance score, the less functional impact a particular amino acid substitution is expected to have.

DISCUSSION

There are many animal species that live on the Tibetan plateau. The plateau consists of 25% of Mainland China, with an average altitude exceeding 4,500 m (Thompson et al., 2000; Wei et al., 2016). Up to now, genomic studies indicated several significant pathways and functional categories including energy metabolism and oxygen transmission, DNA repair, and ATPase production, in response to hypoxia at the Tibetan plateau (Qiu et al., 2012; Ge et al., 2013; Li et al., 2013; Qu et al., 2013; Xiang et al., 2013). Moreover, SNP data have been obtained from population surveys to find the mechanism underlying plateau adaptation (Wei et al., 2016) and have successfully identified candidate genes *EGLN1*, *PPARA*, and *EPAS1* as transcription factors with a pivotal role in adaptation (Buroker et al., 2012; Xiang et al., 2013; Gou et al., 2014; Wang et al., 2014).

In our previous exome sequencing of 330 cashmere goats, we identified a genomic region containing a desmosome gene *DSG3* (Desmogelin3) under strong selection. *DSG3* gene is a desmogelin protein-coding gene located in a cluster on goat chromosome 24, consisting of 16 exons.

In this study, the entire region of (44 kbp) *DSG3* was sequenced. Subsequently, the single-nucleotide polymorphism was identified in a large population of indigenous Chinese goats to determine the genetic difference between the lowland and highland goat populations. Our results showed polymorphism in *DSG3* (a) segregated indigenous Chinese goat population in two clusters; (b) a significant positive linear correlation between the elevated altitude and the frequency of a mutant allele; (c) non-synonymous candidate mutations indicating the contribution of gene in high-altitude adaptation of Tibetan goats.

Moreover, our previous studies suggested that the positive directional selections of genes such as *EPAS1* (Endothelial Per-ARNT-Sim (PAS) domain protein 1), *EDNRA* (Endothelin-1 receptor Precursor), *SIRT1* (Sirtuin type 1), and ryanodine receptor 1 (*RYR1*) are linked to hypoxia. *EPAS1*, *RYR1*, *DSG2* (Desmogelin2) are related to cardiomyopathy and *PTPRJ* (Receptor-type tyrosine-protein phosphatase eta), *FUT1* (fucosyltransferase1), *HEG1* (heart development protein with EGF-like domains 1), *PTPRZ1* (tyrosine phosphatase receptor-type Z polypeptide 1), *SIGLEC1* (Sialic acid-binding Ig-like lectin 1 sialoadhesin), *NPC1L1* (Niemann-Pick disease-type C1 gene-like 1), and *NES* (Nestin) are connected with the cardiovascular system, whereas *DSG3* is important in maintaining the normal structure and function of hairs (Song et al., 2016). Also, some studies suggested that the *DSG3* gene is associated with skin development (Koch et al., 1998; Song et al., 2016).

A high phenotypic variation exists in the Cashmere goat introduced by natural and artificial selection for over 10,000 years, in line with human demand (Naderi et al., 2008; Wang et al., 2016). Owing to evolutionary and artificial selection, the goat has been adapting to different harsh environmental conditions (Song et al., 2016). Multiple genes are described in the significant role of hair and fleece growth in cashmere goats, such as the *POU1F1* and *PRL* genes associated with cashmere yield, diameter, and length, whereas the *LHX2* and *LIM* genes regulate the generation of hair. The *FGF5* gene is a key regulator that controls hair length and the *FGF9* gene to promote the regeneration of hair follicle after wounding. The *WNT2* gene is involved in the initiation of hair follicle (Wang et al., 2016). Furthermore, the *DSG1* gene is involved in the communication of hair follicle cell and morphogenesis of hair follicle (Hanakawa et al., 2004; Dong et al., 2013). The *DSG4* gene affects wool traits and is responsible for coat color in Cashmere goats (E et al., 2016). Furthermore, the *DSG3* gene may have linked with additional fiber growth in the cashmere goats. Other studies have reported that the *DSG1*, *DSG3*, and *DSG4* genes were expressed in keratinocytes of the basal and immediate suprabasal cell layers (Koch et al., 1998; Bazzi et al., 2006; Dusek et al., 2006). According to Nazari-Ghadikolaei et al. (2018), keratinocytes produce keratin, the main protein for hair, nail, and skin synthesis (Nazari-Ghadikolaei et al., 2018).

The biological function of *DSG3* is keratinocyte cell-to-cell adhesion in the basal and suprabasal layers of stratified squamous epithelia. *DSG3* has been shown to be the self-antigen for autoantibodies in the pemphigus vulgaris (PV) disease and is known as PV antigen (PVA) (Koch et al., 1997). In past studies of mice and humans, *DSG3* was predominantly expressed in the outer layer of skin tissues such as basal and suprabasal cell layers of epidermis (Kljuic et al., 2003), whereas other members of the desmogelin gene family *DSG1* are only expressed in suprabasal cell layers and those of *DSG2* in the basal cell layer (Koch et al., 1997). *DSG3* provide desmogeling compensations with *DSG1* in pemphigus foliaceus (PF), which is a potentially fatal autoimmune blistering skin disease in which autoantibodies against *DSG3* and *DSG1* cause loss of keratinocyte cell adhesion (Payne et al., 2005).

Similarly, desmogelin 3 (*DSG3*), a *SLC24A5* gene, is involved in lighter skin pigmentation in Europeans, showing strong signals of positive selection in the high-altitude populations (Huerta-Sanchez et al., 2013). Furthermore, desmogelin genes (*DSG1-4*) are expressed in the epidermis and myocardium tissue and their disruption causes some autoimmune diseases that affect the skin, heart, and mucous membranes in humans

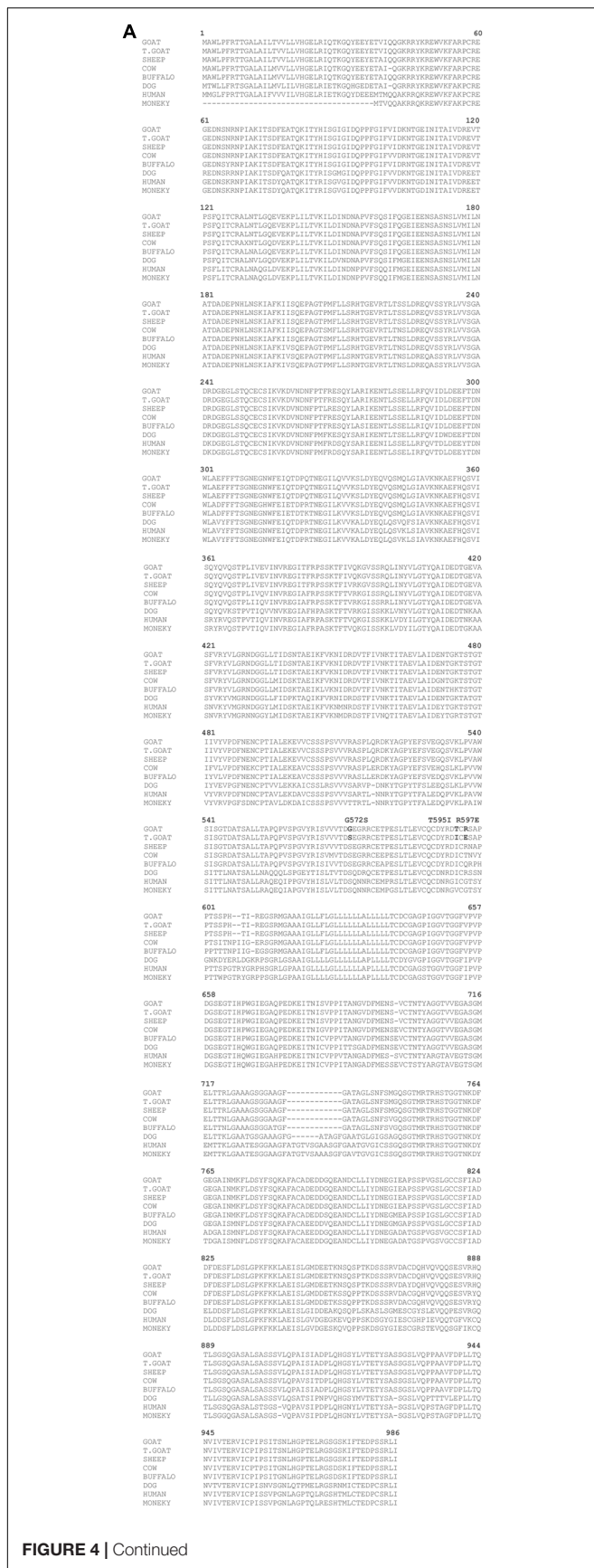


FIGURE 4 | Continued

(Garrod and Chidgey, 2008; Thomason et al., 2010). Desmoglein3 (*DSG3*) has been identified as one of the autoantigens in an autoimmune blistering skin disease called pemphigus vulgaris (PV) (Stanley and Amagai, 2006; Hartlieb et al., 2013). In this disease, circulating autoantibodies targeting *DSG3* induce loss of cell cohesion within the epidermis and mucous membranes. *DSG3* was considered to be the new gene marker for terminal differentiation and was found to be a highly upregulated gene during early phase acute hypobaric hypoxia in the rats (Sharma et al., 2015). There were a few studies related to the functional role of *DSG3*; however, its clear function has not been established yet. It has been suggested that *DSG3* may involve more than cell-cell adhesion (Tsang et al., 2010). The expression of the *DSG3* gene mostly occurred in the stratified epithelial tissues (Stanley and Amagai, 2006). An epithelial tissue plays an essential role in the diseases of altitude because these tissue cells are the lining of the alveoli of lungs (air sacs) and the blood vessels (endothelium). These tissues have no blood vessels and thus must receive oxygen from the blood vessels in the adjacent connective tissue. For this reason, the tissues are particularly susceptible to damage from low oxygen at high altitude. High-altitude hypoxia impacts structures of vital cells such as sodium and potassium pumps and transcription of the genes slows the sodium and potassium pumps in the alveoli of the lungs. Decreasing the number of these pumps contributes to the accumulation of water in the alveoli and causes high-altitude pulmonary edema (Litch, 2006). Several candidate genes have a key role in controlling the complex cellular processes, energy metabolism, oxygen respiration chain, and reform adenosine triphosphate (ATP) in the cell tissues (Niu et al., 2017). During the evolutionary process, the genes functioning within the oxygen respirations chain have been considered under positive selections. These genes are important for adaptation to high-altitude environments (Xu et al., 2007). Our current study has also shown the importance of Desmosomes genes in high-altitude hypoxia adaptation which has been supported by the SNP data validation analysis.

To our knowledge, there are no research publications related to polymorphism of the *DSG3* gene and its role for adaptation to hypoxia at high altitude. In this study, twenty-seven SNPs sites were found from the whole sequence of the *DSG3* gene. The allele frequency of variants was observed to be very low in lowland than in highland goat populations. These results have indicated that allele frequencies of variants might be associated with evolution, selection, genetic variation, and ecological conditions (Li et al., 2004a; Wang et al., 2011). Pearson correlation analysis of SNP1, SNP2, SNP3, SNP4, and SNP5 showed a significant relationship between allele frequency and elevation of altitude. These results are in line with the *EPAS1* gene in the Cashmere goat and Tibetan dog (Fan et al., 2015; Song et al., 2016).

Similarly, the global F_{ST} value varied among the Chinese indigenous goat population at all the SNPs sites. The pairwise genetic distances were an ideal index to measure the polymorphism between the groups. Some exons and introns were excluded from further analysis due to a lower pairwise F_{ST} value. Three non-synonymous and two synonymous mutations have shown large polymorphisms and pairwise genetic distance

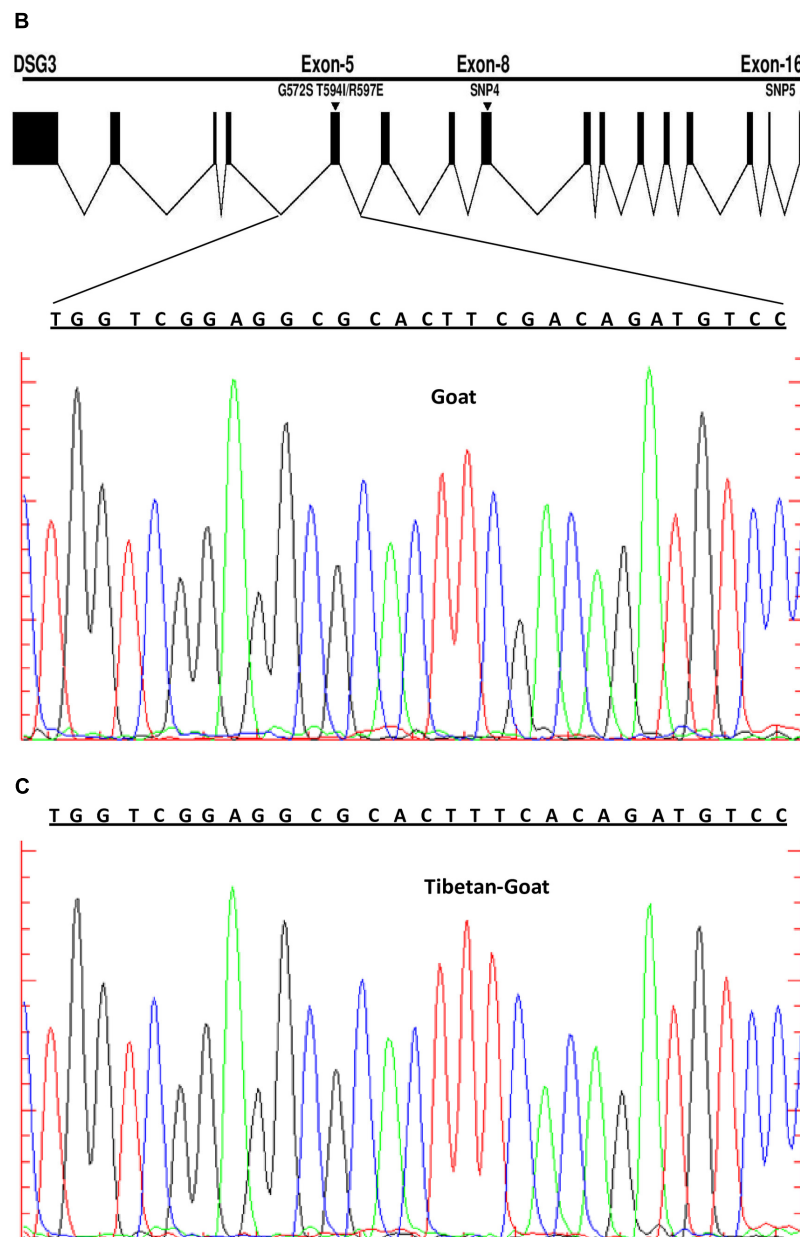


Figure 4 | (A) Protein sequence alignment among different mammalian aligned by cluster omega (Sievers et al., 2011) predicated non-synonymous mutation in Tibetan goat, a conserved site or polymorphism. The dashes indicate deletions. **(B)** Sanger sequence of blue and black region showed nucleotide position changed in lowland goat, blue indicated nucleotide C, while black indicated nucleotide G. **(C)** Sanger sequence of red and blue region indicated nucleotide position changed in Tibetan goat, red indicated nucleotide T, while Blue indicated nucleotide C.

in the *DSG3* gene. Furthermore, this segregated the goat population in distinct branches, as discussed in previously reported studies (Xiang-Long and Valentini, 2004; Di et al., 2011; Ling et al., 2012). Linkage disequilibrium and haplotype analysis showed that the non-synonymous SNPs are highly linked with synonymous SNPs and had higher haplotype frequency variations in the genome region of *DSG3* in highland goat populations (Nordborg and Tavaré, 2002; Conrad et al., 2006; Li et al., 2006).

In this study, three non-synonymous mutations are found from the *DSG3* gene, out of which one mutation G597S is similar to the mutation G305S at *EPAS1* and G14S at *HBB* (Gou et al., 2014; Fan et al., 2015) in the dog. The G305S is an important mutation with a protein functional change in the PAS domain and has a major effect on the blood flow resistance (Adzhubei et al., 2010; Gou et al., 2014). Moreover, it has been demonstrated as a strongly conserved amino acid site at *EPAS1* in the dog (Fan et al., 2015). Therefore, the non-synonymous

mutation G597S is likely to affect the function of the *DSG3* genes (Figures 4A–C) and suggests a role in high-altitude hypoxic adaptation.

Gene mutations and polymorphism are important to understand selection, adaptation, and biological evolutionary processes and to identify the genes related to genetic disease and particular traits (Akey et al., 2002; Nielsen, 2005). The single-nucleotide polymorphism (SNPs) variations in the exonic region of DNA sequences have shown many evolutionary changes. Although the substitution of non-synonymous mutations changed the coding sequence and altered the function of the protein, whereas the synonymous SNP did not change the coding sequence, it may affect the timing of cotranslational protein folding (Sun et al., 2013). In our studies, we have explored five significant SNPs in the most important skin gene *DSG3*, including three non-synonymous and two synonymous substitutes. They have shown a significant positive linear correlation between mutant allele frequency and elevated altitude in the Chinese indigenous goat population.

Future investigations of the *DSG3* gene in goats should focus on the expression level of the gene associated with mutations and altitude. This study has provided the information on the potential function of an allele of *DSG3* in high adaptation and strengthens the understanding of adaptation of the Tibetan goat to the extremely harsh environment of high altitude.

CONCLUSION

Our research has, for the first time, shown the role of the Desmosomes gene for the high-altitude hypoxia adaptation and abundant genetic diversity between goat populations.

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Ai, H., Yang, B., Li, J., Xie, X., Chen, H., and Ren, J. (2014). Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. *BMC Genomics* 15:834. doi: 10.1186/1471-2164-15-834
- Akey, J. M., Zhang, G., Zhang, K., Jin, L., and Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12, 1805–1814. doi: 10.1101/gr.631202
- Bandelt, H. J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036
- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Bazzi, H., Getz, A., Mahoney, M. G., Ishida-Yamamoto, A., Langbein, L., Wahl, J. K. III., et al. (2006). Desmoglein 4 is expressed in highly differentiated keratinocytes and trichocytes in human epidermis and hair follicle. *Differentiation* 74, 129–140. doi: 10.1111/j.1432-0436.2006.00061.x
- Buroker, N. E., Ning, X. H., Zhou, Z. N., Li, K., Cen, W. J., Wu, X. F., et al. (2012). EPAS1 and EGLN1 associations with high altitude sickness in han and tibetan chinese at the qinghai-tibetan plateau. *Blood Cells Mol. Dis.* 49, 67–73. doi: 10.1016/j.bcmd.2012.04.004
- Cai, Q., Qian, X., Lang, Y., Luo, Y., Xu, J., Pan, S., et al. (2013). Genome sequence of ground tit *Pseudopodoces humilis* and its adaptation to high altitude. *Genome Biol.* 14:R29. doi: 10.1186/gb-2013-14-3-r29
- Conrad, D. F., Jakobsson, M., Coop, G., Wen, X., Wall, J. D., Rosenberg, N. A., et al. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* 38, 1251–1260. doi: 10.1038/ng1911
- Crawford, D. C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M. J., Nickerson, D. A., et al. (2004). Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat. Genet.* 36, 700–706. doi: 10.1038/ng1376
- Delva, E., Tucker, D. K., and Kowalczyk, A. P. (2009). The desmosome. *Cold Spring Harb. Perspect. Biol.* 1:a002543. doi: 10.1101/cshperspect.a002543
- Di, R., Vahidi, S. M., Ma, Y. H., He, X. H., Zhao, Q. J., Han, J. L., et al. (2011). Microsatellite analysis revealed genetic diversity and population structure among chinese cashmere goats. *Anim. Genet.* 42, 428–431. doi: 10.1111/j.1365-2052.2010.02072.x
- Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., et al. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* 31, 135–141. doi: 10.1038/nbt.2478

We found that three non-synonymous candidate mutations (R597E, T595I, and G572S) and two synonymous substitutions significantly segregated these goat populations. Moreover, these results indicated the contribution of *DSG3* in high-altitude adaptation and provided new insights for Tibetan cashmere goats.

AUTHOR CONTRIBUTIONS

CK and LJ wrote the manuscript. XH, QZ, and YP contributed to sample collections and designing the manuscripts. CK and SS has performed the data analysis. YM and LJ conceived the study design. YM, KKM, and AAK interpreted the results and revised the manuscript. All the authors read and approved the final manuscript.

FUNDING

This research was funded by the National Natural Science Foundation of China (Nos. 31472064 and 31601910), the Special Fund for Agro-Scientific Research in the Public Interest (201303059), and the earmarked fund for Modern Agro-industry Technology Research System (CARS-40-01). LJ was supported by the Elite Youth Program in the Chinese Academy of Agricultural Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00553/full#supplementary-material>

- Dusek, R. L., Getsios, S., Chen, F., Park, J. K., Amargo, E. V., Cryns, V. L., et al. (2006). The differentiation-dependent desmosomal cadherin desmoglein 1 is a novel caspase-3 target that regulates apoptosis in keratinocytes. *J. Biol. Chem.* 281, 3614–3624. doi: 10.1074/jbc.M508258200
- E, G. X., Zhao, Y. J., Ma, Y. H., Cao, G. L., He, J. N., Na, R. S., et al. (2016). Desmoglein 4 diversity and correlation analysis with coat color in goat. *Genet. Mol. Res.* 15:15017814. doi: 10.4238/gmr.15017814
- Fan, R., Liu, F., Wu, H., Wu, S., Zhu, C., Li, Y., et al. (2015). A positive correlation between elevated altitude and frequency of mutant alleles at the EPAS1 and HBB loci in chinese indigenous dogs. *J. Genet. Genomics* 42, 173–177. doi: 10.1016/j.jgg.2015.02.006
- Garrod, D., and Chidgey, M. (2008). Desmosome structure, composition and function. *Biochim. Biophys. Acta* 1778, 572–587. doi: 10.1016/j.bbamem.2007.07.014
- Ge, R. L., Cai, Q., Shen, Y. Y., San, A., Ma, L., Zhang, Y., et al. (2013). Draft genome sequence of the Tibetan antelope. *Nat. Commun.* 4:1858. doi: 10.1038/ncomms2860
- Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., et al. (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* 24, 1308–1315. doi: 10.1101/gr.171876.113
- Hanakawa, Y., Li, H., Lin, C., Stanley, J. R., and Cotsarelis, G. (2004). Desmogleins 1 and 3 in the companion layer anchor mouse anagen hair to the follicle. *J. Invest. Dermatol.* 123, 817–822. doi: 10.1111/j.0022-202X.2004.23479.x
- Hartlieb, E., Kempf, B., Partilla, M., Vigh, B., Spindler, V., and Waschke, J. (2013). Desmoglein 2 is less important than desmoglein 3 for keratinocyte cohesion. *PLoS One* 8:e53739. doi: 10.1371/journal.pone.0053739
- Huerta-Sanchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., et al. (2013). Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Mol. Biol. Evol.* 30, 1877–1888. doi: 10.1093/molbev/mst089
- Kljuic, A., Bazzi, H., Sundberg, J. P., Martinez-Mir, A., O'Shaughnessy, R., Mahoney, M. G., et al. (2003). Desmoglein 4 in hair follicle differentiation and epidermal adhesion. *Cell* 113, 249–260. doi: 10.1016/S0092-8674(03)00273-3
- Koch, P. J., Mahoney, M. G., Cotsarelis, G., Rothenberger, K., Lavker, R. M., and Stanley, J. R. (1998). Desmoglein 3 anchors telogen hair in the follicle. *J. Cell Sci.* 111(Pt 17), 2529–2537.
- Koch, P. J., Mahoney, M. G., Ishikawa, H., Pulkkinen, L., Uitto, J., Shultz, L., et al. (1997). Targeted disruption of the pemphigus vulgaris antigen (Desmoglein 3) gene in mice causes loss of keratinocyte cell adhesion with a phenotype similar to pemphigus vulgaris. *J. Cell Biol.* 137:1091.
- Li, C., Wang, Z., Liu, B., Yang, S., Zhu, Z., Fan, B., et al. (2004a). Evaluation of the genetic relationship among ten chinese indigenous pig breeds with twenty-six microsatellite markers. *Asian-Australas J Anim Sci* 17, 441–444. doi: 10.5713/ajas.2004.441
- Li, M. H., Li, K., and Zhao, S. H. (2004b). Diversity of chinese indigenous goat breeds: a conservation perspective - a review. *Asian-Australas J. Anim. Sci.* 17, 726–732. doi: 10.5713/ajas.2004.726
- Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi: 10.1038/ng.2811
- Li, X.-L., Wu, Z.-L., Liu, Z.-Z., Gong, Y.-F., Zhou, R.-Y., and Zheng, G.-R. (2006). SNP identification and analysis in part of intron 2 of goat MSTN gene and variation within and among species. *J. Hered.* 97, 285–289. doi: 10.1093/jhered/esj026
- Li, Y., Wu, D. D., Boyko, A. R., Wang, G. D., Wu, S. F., Irwin, D. M., et al. (2014). Population variation revealed high-altitude adaptation of tibetan mastiffs. *Mol. Biol. Evol.* 31, 1200–1205. doi: 10.1093/molbev/msu070
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Ling, Y. H., Zhang, X. D., Yao, N., Ding, J. P., Chen, H. Q., Zhang, Z. J., et al. (2012). Genetic differentiation of chinese indigenous meat goats ascertained using microsatellite information. *Asian-Australas J. Anim. Sci.* 25, 177–182. doi: 10.5713/ajas.2011.11308
- Litch, J. A. (2006). Going higher: oxygen, man, and mountains, 5th Edition. *Wilderness Environ. Med.* 17:e16–e17. doi: 10.1016/S1080-6032(06)70891-2
- Naderi, S., Rezaei, H. R., Pompanon, F., Blum, M. G., Negrini, R., Naghash, H. R., et al. (2008). The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17659–17664. doi: 10.1073/pnas.0804782105
- Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S. R., Staiger, E. A., Rashidi, A., and Huson, H. J. (2018). Genome-wide association studies identify candidate genes for coat color and mohair traits in the iranian markhoz goat. *Front. Genet.* 9:105. doi: 10.3389/fgene.2018.00105
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218. doi: 10.1146/annurev.genet.39.073003.112420
- Niu, L., Chen, X., Xiao, P., Zhao, Q., Zhou, J., Hu, J., et al. (2017). Detecting signatures of selection within the tibetan sheep mitochondrial genome. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 28, 801–809. doi: 10.1080/24701394.2016.1192614
- Nordborg, M., and Tavaré, S. (2002). Linkage disequilibrium: what history has to tell us. *Trends Genet.* 18, 83–90. doi: 10.1016/S0168-9525(02)00255-X
- Payne, A. S., Ishii, K., Kacir, S., Lin, C., Li, H., Hanakawa, Y., et al. (2005). Genetic and functional characterization of human pemphigus vulgaris monoclonal autoantibodies isolated by phage display. *J. Clin. Invest.* 115, 888–899. doi: 10.1172/JCI24185
- Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., et al. (2012). The yak genome and adaptation to life at high altitude. *Nat. Genet.* 44, 946–949. doi: 10.1038/ng.2343
- Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B., et al. (2013). Ground tit genome reveals avian adaptation to living at high altitudes in the tibetan plateau. *Nat. Commun.* 4:2071. doi: 10.1038/ncomms3071
- Sharma, P., Bansal, A., and Sharma, P. C. (2015). RNA-seq-based transcriptome profiling reveals differential gene expression in the lungs of sprague-dawley rats during early-phase acute hypobaric hypoxia. *Mol. Genet. Genomics* 290, 2225–2240. doi: 10.1007/s00438-015-1064-0
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75
- Song, S., Yao, N., Yang, M., Liu, X., Dong, K., Zhao, Q., et al. (2016). Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the tibetan cashmere goat (*Capra hircus*). *BMC Genomics* 17:122. doi: 10.1186/s12864-016-2449-0
- Stanley, J. R., and Amagai, M. (2006). Pemphigus, bullous impetigo, and the staphylococcal scalded-skin syndrome. *N. Engl. J. Med.* 355, 1800–1810. doi: 10.1056/NEJMra061111
- Stephens, M., Smith, N. J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989. doi: 10.1086/319501
- Sun, J., Zhong, H., Chen, S. Y., Yao, Y. G., and Liu, Y. P. (2013). Association between MT-CO3 haplotypes and high-altitude adaptation in tibetan chicken. *Gene* 529, 131–137. doi: 10.1016/j.gene.2013.06.075
- Thomason, H. A., Scothern, A., McHarg, S., and Garrod, D. R. (2010). Desmosomes: adhesive strength and signalling in health and disease. *Biochem. J.* 429, 419–433. doi: 10.1042/bj20100567
- Thompson, L. G., Yao, T., Mosley-Thompson, E., Davis, M., Henderson, K., and Lin, P.-N. (2000). A high-resolution millennial record of the south asian monsoon from himalayan ice cores. *Science* 289, 1916–1919. doi: 10.1126/science.289.5486.1916
- Tsang, S. M., Liu, L., Teh, M. T., Wheeler, A., Grose, R., Hart, I. R., et al. (2010). Desmoglein 3, via an interaction with E-cadherin, is associated with activation of Src. *PLoS One* 5:e14211. doi: 10.1371/journal.pone.0014211
- Wang, G. D., Fan, R. X., Zhai, W., Liu, F., Wang, L., Zhong, L., et al. (2014). Genetic convergence in the adaptation of dogs and humans to the high-altitude

- environment of the tibetan plateau. *Genome Biol. Evol.* 6, 2122–2128. doi: 10.1093/gbe/evu162
- Wang, X., Liu, J., Zhou, G., Guo, J., Yan, H., Niu, Y., et al. (2016). Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci. Rep.* 6:38932. doi: 10.1038/srep38932
- Wang, Y., Wang, J., Zi, X. D., Huatai, C. R., Ouyang, X., and Liu, L. S. (2011). Genetic diversity of tibetan goats of plateau type using microsatellite markers. *Arch. Anim. Breed* 54, 188–197. doi: 10.5194/aab-54-188-2011
- Wei, C., Wang, H., Liu, G., Zhao, F., Kijas, J. W., Ma, Y., et al. (2016). Genome-wide analysis reveals adaptation to high altitudes in tibetan sheep. *Sci. Rep.* 6:26770. doi: 10.1038/srep26770
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641
- Xiang, K., Ouzhuluobu, Peng, Y., Yang, Z., Zhang, X., Cui, C., et al. (2013). Identification of a tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Mol. Biol. Evol.* 30, 1889–1898. doi: 10.1093/molbev/mst090
- Xiang-Long, L., and Valentini, A. (2004). Genetic diversity of chinese indigenous goat breeds based on microsatellite markers. *J. Anim. Breed. Genet.* 121, 350–355. doi: 10.1111/j.1439-0388.2004.00465.x
- Xu, S., Luosang, J., Hua, S., He, J., Ciren, A., Wang, W., et al. (2007). High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *J. Genet. Genomics* 34, 720–729. doi: 10.1016/s1673-8527(07)60081-2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kumar, Song, Jiang, He, Zhao, Pu, Malhi, Kamboh and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Dissecting the Genomic Architecture of Resistance to *Eimeria maxima* Parasitism in the Chicken

Kay Boulton^{1*}, Matthew J. Nolan², Zhiguang Wu¹, Valentina Riggio¹, Oswald Matika¹, Kimberley Harman², Paul M. Hocking^{1†}, Nat Bumstead^{3†}, Pat Hesketh³, Andrew Archer³, Stephen C. Bishop^{1†}, Pete Kaiser^{1†}, Fiona M. Tomley², David A. Hume^{1,4}, Adrian L. Smith^{3,5}, Damer P. Blake² and Androniki Psifidi^{1,2,6}

¹ The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, United Kingdom,

² Department of Pathobiology and Population Sciences, Royal Veterinary College, University of London, London, United Kingdom, ³ Enteric Immunology Group and Genetics and Genomics Group, Pirbright Institute, Woking,

United Kingdom, ⁴ Mater Research Institute, The University of Queensland, Brisbane, St. Lucia, QLD, Australia, ⁵ Department

of Zoology, Sir Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom, ⁶ Department

of Clinical Sciences and Services, Royal Veterinary College, University of London, Hatfield, United Kingdom

OPEN ACCESS

Edited by:

John Anthony Hammond,
Pirbright Institute, United Kingdom

Reviewed by:

Emanuel Heitlinger,
Leibniz Institute for Zoo and Wildlife
Research (LG), Germany
Michael Kogut,
Agricultural Research Service (USDA),
United States

*Correspondence:

Kay Boulton
kay.boulton@roslin.ed.ac.uk

†Deceased

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 31 May 2018

Accepted: 22 October 2018

Published: 26 November 2018

Citation:

Boulton K, Nolan MJ, Wu Z,
Riggio V, Matika O, Harman K,
Hocking PM, Bumstead N, Hesketh P,
Archer A, Bishop SC, Kaiser P,
Tomley FM, Hume DA, Smith AL,
Blake DP and Psifidi A (2018)
Dissecting the Genomic Architecture
of Resistance to *Eimeria maxima*
Parasitism in the Chicken.
Front. Genet. 9:528.
doi: 10.3389/fgene.2018.00528

Coccidiosis in poultry, caused by protozoan parasites of the genus *Eimeria*, is an intestinal disease with substantial economic impact. With the use of anticoccidial drugs under public and political pressure, and the comparatively higher cost of live-attenuated vaccines, an attractive complementary strategy for control is to breed chickens with increased resistance to *Eimeria* parasitism. Prior infection with *Eimeria maxima* leads to complete immunity against challenge with homologous strains, but only partial resistance to challenge with antigenically diverse heterologous strains. We investigate the genetic architecture of avian resistance to *E. maxima* primary infection and heterologous strain secondary challenge using White Leghorn populations of derived inbred lines, C.B12 and 15I, known to differ in susceptibility to the parasite. An intercross population was infected with *E. maxima* Houghton (H) strain, followed 3 weeks later by *E. maxima* Weybridge (W) strain challenge, while a backcross population received a single *E. maxima* W infection. The phenotypes measured were parasite replication (counting fecal oocyst output or qPCR for parasite numbers in intestinal tissue), intestinal lesion score (gross pathology, scale 0–4), and for the backcross only, serum interleukin-10 (IL-10) levels. Birds were genotyped using a high density genome-wide DNA array (600K, Affymetrix). Genome-wide association study located associations on chromosomes 1, 2, 3, and 5 following primary infection in the backcross population, and a suggestive association on chromosome 1 following heterologous *E. maxima* W challenge in the intercross population. This mapped several megabases away from the quantitative trait locus (QTL) linked to the backcross primary W strain infection, suggesting different underlying mechanisms for the primary- and heterologous secondary- responses. Underlying pathways for those genes located in the respective QTL for resistance to primary infection and protection against heterologous challenge were related mainly to immune response, with IL-10 signaling in the backcross primary infection being the most significant. Additionally, the identified markers associated

with IL-10 levels exhibited significant additive genetic variance. We suggest this is a phenotype of interest to the outcome of challenge, being scalable in live birds and negating the requirement for single-bird cages, fecal oocyst counts, or slaughter for sampling (qPCR).

Keywords: intercross, backcross, *Eimeria maxima*, QTL, resistance, interleukin-10, oocyst output

INTRODUCTION

Coccidiosis is an intestinal disease caused by intracellular protozoan parasites of the genus *Eimeria* (Shirley et al., 2005). The control of coccidiosis is a challenge to the international poultry industry, with economic losses estimated at USD 3 billion annually (Dalloul and Lillehoj, 2006). Current control of coccidiosis relies on the prophylactic use of anticoccidial drugs, or vaccination with formulations of live wild-type or attenuated parasites (Crouch et al., 2003; McDonald and Shirley, 2009). However, use of some anticoccidial drugs has been curtailed by legislation, while the limited production capacity and costs of live attenuated vaccines compromise their utility in broiler flocks (Hong et al., 2006). Thus, there is a need for complementary strategies to control coccidiosis in poultry. A promising approach would be to breed chickens for increased genetic resistance and increased vaccine response to *Eimeria* parasitism since there is evidence for relevant host genetic variation (Johnson et al., 1986; Bumstead and Millard, 1992).

Coccidiosis in poultry is caused by seven distinct *Eimeria* species (Reid et al., 2014), with *Eimeria maxima* being one of the most common causes of coccidiosis in commercial broilers. Immunity introduced by primary infection (vaccination) against *E. maxima* is commonly strain-specific, with immune escape contributing to sub-clinical coccidiosis symptoms that include decreased feed conversion efficiency, marked weight loss and low performance (Fitz-Coy, 1992; Blake et al., 2005).

Johnson et al. (1986) demonstrated variance in coccidiosis susceptibility in chickens as a prerequisite to selective breeding for resistance. A subsequent study using several inbred White Leghorn lines established variance for benchmark phenotypes when chickens were infected with controlled doses of *Eimeria* spp. (Bumstead and Millard, 1987; Bumstead and Millard, 1992). The between-line variation observed in oocyst production by the different lines was not correlated with weight loss or mortality, indicating that within-trait observations were a result of effect accommodation rather than parasite restriction. The greatest differences in parasite replication (PR) were between lines 15I and C major histocompatibility complex (MHC) haplotype B12 (C.B12) chickens that produced relatively high and low numbers of oocysts, respectively (Bumstead and Millard, 1987; Smith et al., 2002). Most notably, primary infection with the Houghton or Weybridge reference *E. maxima* strains induce 100% protection against secondary homologous challenge in 15I and C line chickens (Smith et al., 2002). However, the outcome of heterologous challenge varied by parasite strain and host genotype combination (Smith et al., 2002; Blake et al., 2004, 2005). Regardless of the substantial financial losses to industry caused by coccidiosis, few studies have attempted to identify

quantitative trait loci (QTL) for resistance to *E. maxima* infection and there are no relevant studies on the genetics of heterologous secondary challenge response.

The present study extends previous work in inbred chicken lines to determine the genetic architecture of *E. maxima* resistance, i.e., lack of PR, and protection against secondary challenge with a heterologous *E. maxima* strain. First, an F2 intercross of inbred White Leghorn chicken lines C.B12 × 15I were initially infected with *E. maxima* H, followed 3 weeks later by challenge with *E. maxima* W to investigate response to challenge with the heterologous strain. Fecal oocyst output was counted to determine severity of challenge. Second, a backcross population from the same two inbred lines [(C.B12 × 15I) × C.B12] was infected with *E. maxima* W to study primary resistance to parasitism. Three phenotypes were determined for these birds following infection: PR by qPCR for parasite numbers in intestinal tissue, intestinal lesion score (LS) (gross pathology, scale 0–4) and levels of serum interleukin-10 (IL-10), a novel biomarker, found to be positively correlated with the pathology trait in chickens infected with *E. tenella* (Wu et al., 2016; Boulton et al., 2018). All birds were then genotyped using a 600K Affymetrix® Axiom® HD array (Kranis et al., 2013), enabling genome-wide association studies (GWASs), followed by pathway analysis to identify candidate genomic regions, pathways, networks and genes for resistance to *E. maxima* primary infection and effective responses to challenge with a heterologous strain.

MATERIALS AND METHODS

Ethics Statement

These trials were conducted under Home Office Project Licence in accordance with Home Office regulations under the Animals (Scientific Procedures) Act 1986 and the guidelines set down by the Institute for Animal Health and RVC Animal Welfare and Ethical Review Bodies.

Parasites

The *E. maxima* Houghton (H) and Weybridge (W) strains were used throughout these studies (Norton and Hein, 1976). Routine parasite passage, sporulation, and dose preparation were undertaken as described previously (Eckert et al., 1995) using specific pathogen free Light Sussex or Lohman LSL chickens. Oocysts were used within 1 month of harvest.

Animals

Inbred chicken lines 15I and C derived from White Leghorn flocks at USDA-ARS Avian Disease and Oncology Laboratory

in East Lansing, MI, United States, were maintained by random mating within the specified-pathogen-free (SPF) flocks at the Pirbright Institute [formerly the Institute for Animal Health (IAH)], United Kingdom since 1962 and 1969, respectively.

F2 intercross birds ($n = 195$) were generated by crossing nine F1 (C.B12 \times 15I) male progeny with 27 unrelated F1 female progeny at the IAH (Compton site). Six birds from each of the two parental lines, 15I and C.B12, were also hatched and kept under the same experimental conditions as F2 (individual cages post-challenge).

To generate the backcross ($n = 214$), 20 F1 (C.B12 \times 15I) male progeny were crossed with 100 unrelated C.B12 line females. The breeding was performed in the SPF Bumstead facility at the Roslin Institute, The University of Edinburgh, United Kingdom. Day old chicks were transported in isolated SPF containment to the Royal Veterinary College poultry barn, University of London, United Kingdom, where the primary infection with *E. maxima* W sporulated oocysts were conducted in floor pens.

***Eimeria maxima* Challenge Experiments**

Intercross Population

F2 intercross ($n = 195$), and 12 parental line birds were initially infected by oral gavage with 100 sporulated oocysts of *E. maxima* H at 25 days of age and moved to individual cages. Feces were collected from each bird on a daily basis during the 5–10 days post-challenge (pi) period following infection. Three weeks later (47 days of age) a secondary challenge was initiated by oral gavage of 250 sporulated oocysts of *E. maxima* W. Feces were again collected from each bird on a daily basis during the 5–10 day post-challenge period.

Backcross Population

At 21 days of age, chickens were inoculated by oral gavage with either 1 ml distilled water (control group, $n = 20$) or 100 sporulated oocysts of *E. maxima* W (infected group, $n = 194$). To avoid cross-infection the control group was housed separately. Birds were euthanised humanely at day 7 pi, coinciding with the peak pathological effects of *E. maxima* (Rothwell et al., 2004), providing the greatest sensitivity for parasite genome detection (Blake et al., 2006). A blood sample from each bird was collected post-mortem *via* aortic rupture into 1.5 ml Sigma-Aldrich (Dorset, United Kingdom) microcentrifuge tubes. Bijou tubes (7 ml SterilinTM) containing 5–10 volumes of room temperature RNAlater[®] (Life Technologies, Carlsbad, CA, United States) were used to store 5.0 cm of intestinal tissue and content from either side of Meckel's diverticulum.

Phenotyping

Individual oocyst output was used to study the outcome of the *E. maxima* H primary infection and secondary heterologous *E. maxima* W challenge in the intercross chicken population. Oocysts were quantified daily (5 to 10 days post-infection and challenge) using a microscope and saturated salt flotation in a McMaster counting chamber (Eckert et al., 1995; Smith et al., 2002). Daily totals were combined to provide a total count for oocyst output per bird for both the primary infection and

secondary challenge. Oocyst counts were log-transformed to approximate normal distribution.

The phenotypes used to study resistance to *E. maxima* W primary infection in the backcross population were relative intestinal *Eimeria* genome copy number (PR, measured using quantitative PCR as parasite genomes per host chicken genome), intestinal LS (pathology, on a scale 0–4), and serum IL-10 level (IL-10). Quantitative real-time PCR targeting the *E. maxima* microneme protein 1 (*EmMIC1*) and *Gallus gallus* β -actin (*actb*) loci was performed using total genomic DNA extracted from a 10 cm length of intestinal tissue centered on Meckel's diverticulum using a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). Briefly, each complete tissue sample was disaggregated using a Qiagen TissueRuptor and an aliquot was processed for extraction of combined host and parasite DNA (see Blake et al., 2006, for full details). A CFX96 Touch[®] Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, United States), was used to amplify each sample in triplicate (Nolan et al., 2015), with an additional Bead-Beater homogenization step prior to buffer ATL treatment (including 1 volume 0.4–0.6 mm glass beads, 3,000 oscillations per minute for 1 min). Intestinal pathology was assessed by the same experienced operator scoring lesions according to Johnson and Reid (1970). A capture ELISA was used to measure IL-10, employing ROS-AV164 and biotinylated ROS-AV163 as capture and detection antibodies, respectively (see Wu et al., 2016, for full details). IL-10 levels and parasite genome numbers were log-transformed to approximate normal distribution.

Phenotypic Correlations

Following log-transformation for PR and IL-10, all backcross phenotypic traits were rescaled to modify the unit of measurement differences. Then, fitting host sex as a fixed effect in a multivariate linear model, phenotypic correlations (r_p) were estimated using ASReml 4.1 (Gilmour et al., 2015).

Genome-Wide Association Studies

Sixty-seven F2 birds exhibiting the most extreme phenotypes, plus the 12 intercross parental line birds and the entire backcross generation were genotyped using the 600K Affymetrix[®] Axiom[®] HD genotyping array (Kranis et al., 2013). Although each data set was analyzed separately, the same GWAS steps were used for both populations. The marker genotype data were subjected to quality control measures using the thresholds: minor allele frequency < 0.02 and call rate $> 90\%$. Deviation from Hardy-Weinberg equilibrium was not considered a reason for excluding markers since these were experimental populations of inbred lines. After quality control 203,845 intercross and 204,072 backcross markers remained and were used, respectively, to generate separate intercross and backcross genomic relationship matrixes (GRMs) to investigate the presence of population stratification. Next, each GRM was converted to a distance matrix that was analyzed with a classical multidimensional scaling using the GenABEL package of R (Aulchenko et al., 2007) to obtain principal components. These analyses revealed three principal components in the intercross population (one for each parental line and one for F2 birds), but no substructure in the backcross.

GWAS for each trait were then conducted using GenABEL based on a mixed model, with the population principal components fitted as a co-variate (intercross population only), sex fitted as a fixed effect in both studies, and GRM fitted as a random polygenic effect to adjust for population sub-structure. In the case of GWAS for heterologous secondary challenge response, the oocyst output following the first challenge was also fitted as a covariate to account for the effect of the first challenge. After Bonferroni correction for multiple testing, significance thresholds were $P \leq 2.45 \times 10^{-7}$ and $P \leq 4.90 \times 10^{-6}$ for genome-wide ($P \leq 0.05$) and suggestive (namely one false positive per genome scan) significant levels, respectively corresponding to $-\log_{10}(P)$ of 6.61 and 5.30. The extent of linkage disequilibrium (LD) between significant markers located on the same chromosome regions was calculated using the r-square statistic of PLINK v1.09 (Purcell et al., 2007).

Effects of the significant markers identified in each GWAS were re-estimated in ASReml 4.1 (Gilmour et al., 2015) by individually fitting the markers as fixed effects in the same model as used for GWAS analyses. Effects were calculated as follows: additive effect, $a = (AA - BB)/2$; dominance effect, $d = AB - ((AA + BB)/2)$, where AA, BB, and AB were the predicted trait values for each genotype class.

All significant markers identified in GWAS for responses to primary infection and secondary *E. maxima* W challenge were mapped to the reference *Gallus gallus domesticus* genome and annotated using the variant effect predictor¹ tool within the Ensembl (genome browser 92) database and the Gal-gal5 assembly². Furthermore, genes located within 100 kb up- and down-stream of the significant markers were annotated using the BioMart data mining tool³ and the Gal-gal5 assembly. This method of annotation enabled all genes located in the vicinity of the identified significant markers to be identified and cataloged.

Re-sequencing Data Analysis

To identify possible protein-coding genes associated with the detected QTL, genomic sequences in the regions of interest from the line 15I and C.B12 chickens were compared. The two parental chicken lines were entirely re-sequenced at 15–20 fold coverage, using pools of 10 individuals per line, performed on an Illumina GAIIx platform using a paired-end protocol (Krämer et al., 2014). Re-sequencing data of the candidate regions (i.e., 1 kb up- and downstream of the candidate gene end sites), for resistance to primary infection and heterologous challenge derived from intercross and backcross analyses, were then extracted and examined separately. Using the Mpileup tool for marker calling (SAMtools v0.1.7; Li et al., 2009), single nucleotide variants (SNVs) between the two parental lines and the reference genome in these regions were detected. These were then annotated using the same variant effect predictor software as above. Information for all SNV [intergenic, intronic, exonic, splicing, 3' and 5' untranslated regions (3' UTR, 5' UTR)] present in the regions of interest were collated. Intergenic, intronic, and

exonic synonymous variants were then filtered out along with SNV that were common in the two parental lines but different from the reference genome. Thus, only sites that were different between the parental lines and had an effect on the coding sequence (nonsense, missense, splicing) or a potential effect on the gene expression (3' UTR and 5' UTR) were retained for further study.

Pathway, Network, and Functional Enrichment Analyses

Identification of potential canonical pathways and networks underlying the candidate genomic regions associated with outcomes of primary infection and heterologous secondary *E. maxima* challenge were performed using the ingenuity pathway analysis (IPA) program⁴. IPA constructs multiple possible upstream regulators, pathways, and networks that serve as hypotheses for the biological mechanism underlying the phenotypes based on a large-scale causal network derived from the Ingenuity Knowledge Base. After correcting for a baseline threshold and calculating statistical significance, the most likely pathways involved are inferred (Krämer et al., 2014). The constructed networks can then be ranked using their IPA score based on the *P*-values obtained using Fisher's exact test [IPA score or *P*-score = $-\log_{10}(P\text{-value})$].

The gene lists for each phenotype were also analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; Dennis et al., 2003). To understand the biological meaning behind these genes, gene ontology (GO) was determined, and functional annotation clustering analysis was performed using the integral *G. gallus* background. The enrichment score (ES) of DAVID is a modified Fisher exact *P*-value calculated by the software, with higher ES reflecting more enriched clusters. An ES > 1 means that the functional category is overrepresented.

RESULTS

Descriptive Statistics

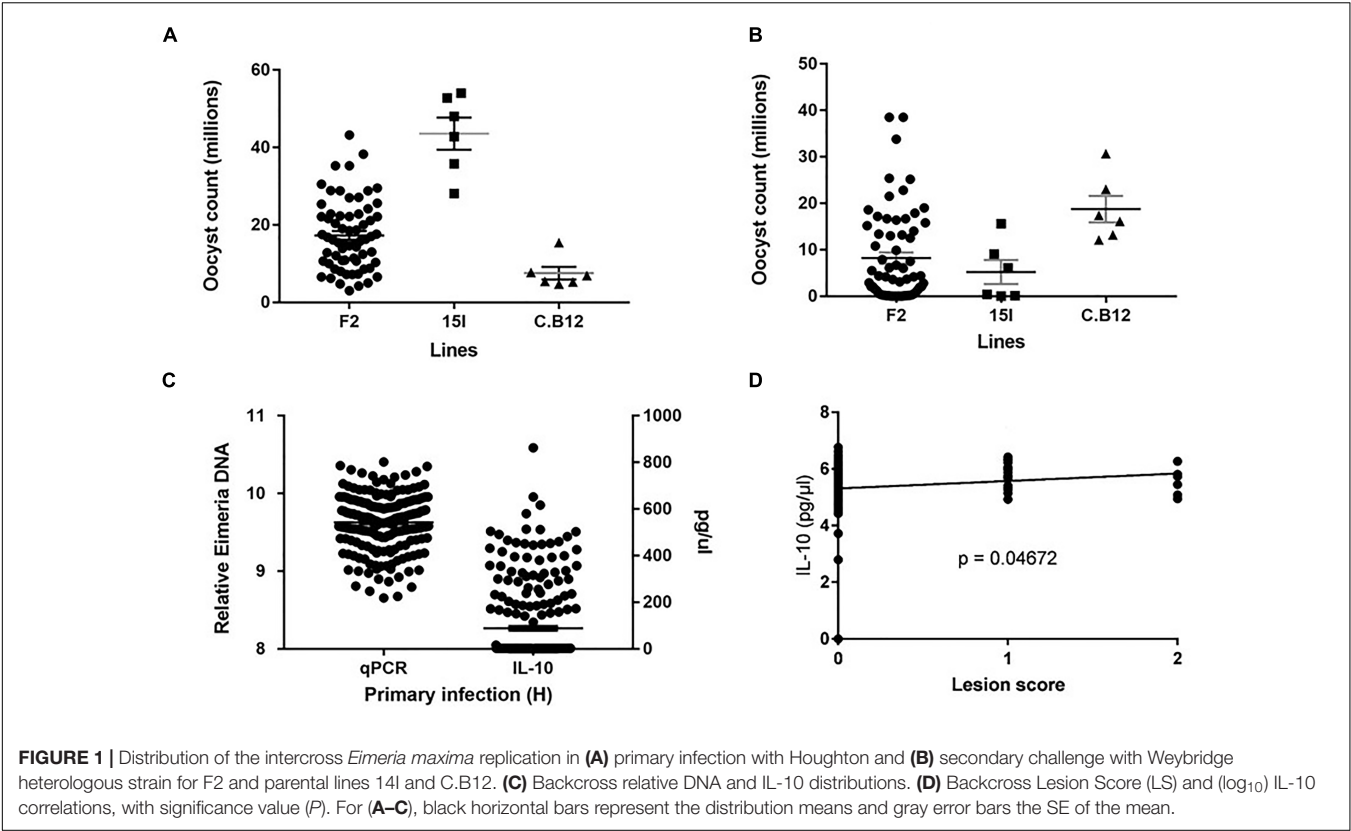
Phenotypic distributions for oocyst counts following primary infection with *E. maxima* H and secondary challenge with *E. maxima* W in the intercross and parental populations along with relative DNA and IL-10 levels in the backcross populations after primary infection with *E. maxima* W are presented in **Figures 1A–C**. After primary infection the pure line C.B12 birds produced fewer *E. maxima* oocyst counts compared to the pure line 15I and F2 birds, with the highest oocyst output recorded in the pure line 15I group. Conversely, inverse findings regarding oocyst output were recorded in the two parental lines following heterologous secondary strain challenge. These results agree with previous findings that show line C.B12 birds develop no cross protection between primary H and secondary W strain challenges, while line 15I birds develop significant cross-protection when infected in this order (Smith et al., 2002; Blake et al., 2005). As expected, for both primary and

¹<http://www.ensembl.org/Tools/VEP>

²https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.4/

³<http://www.ensembl.org/biomart/martview/>

⁴www.ingenuity.com



secondary challenges F2 intercross population oocyst count level was intermediate between those of the two parental lines. Among the backcross chickens, following infection with *E. maxima* W, phenotypic scores for intestinal lesions were low (0–2), however significant variance (*P* = 0.05) was noted (Table 1). Estimated phenotypic correlations between the three measured traits ranged from 0.8 to 0.15, with only the correlation between LS and IL-10 being statistically significant ($r_{LS,IL-10} = 0.15 \pm 0.07$; Figure 1D and Table 1).

Genome-Wide Association Studies Intercross Study

Genome-wide association study analysis for oocyst output following primary infection of the intercross population with *E. maxima* H did not reveal significant associations after

the strict Bonferroni correction. However, an association with markers on chromosome 2, just below the suggestive threshold was reported (results not shown). GWAS analysis following secondary challenge with the heterologous *E. maxima* W strain identified 11 markers on chromosome 1, all having suggestive associations with the trait in the intercross population. These 11 markers belonged to the same LD block (499 bp, $r^2 = 1$; Figure 2 and Table 2). The corresponding Q–Q plot for the GWAS intercross result is found in Figure 2.

The 11 significant markers associated with the outcome of secondary challenge by the heterologous *E. maxima* strain were all located in intronic, upstream, and downstream regions of the *phenylalanine hydroxylase* (*PAH*) gene (Supplementary Table S1). In the 0.5 Mb candidate region for enhanced response to heterologous secondary *E. maxima* challenge only 16 protein coding genes were located (Supplementary Table S2).

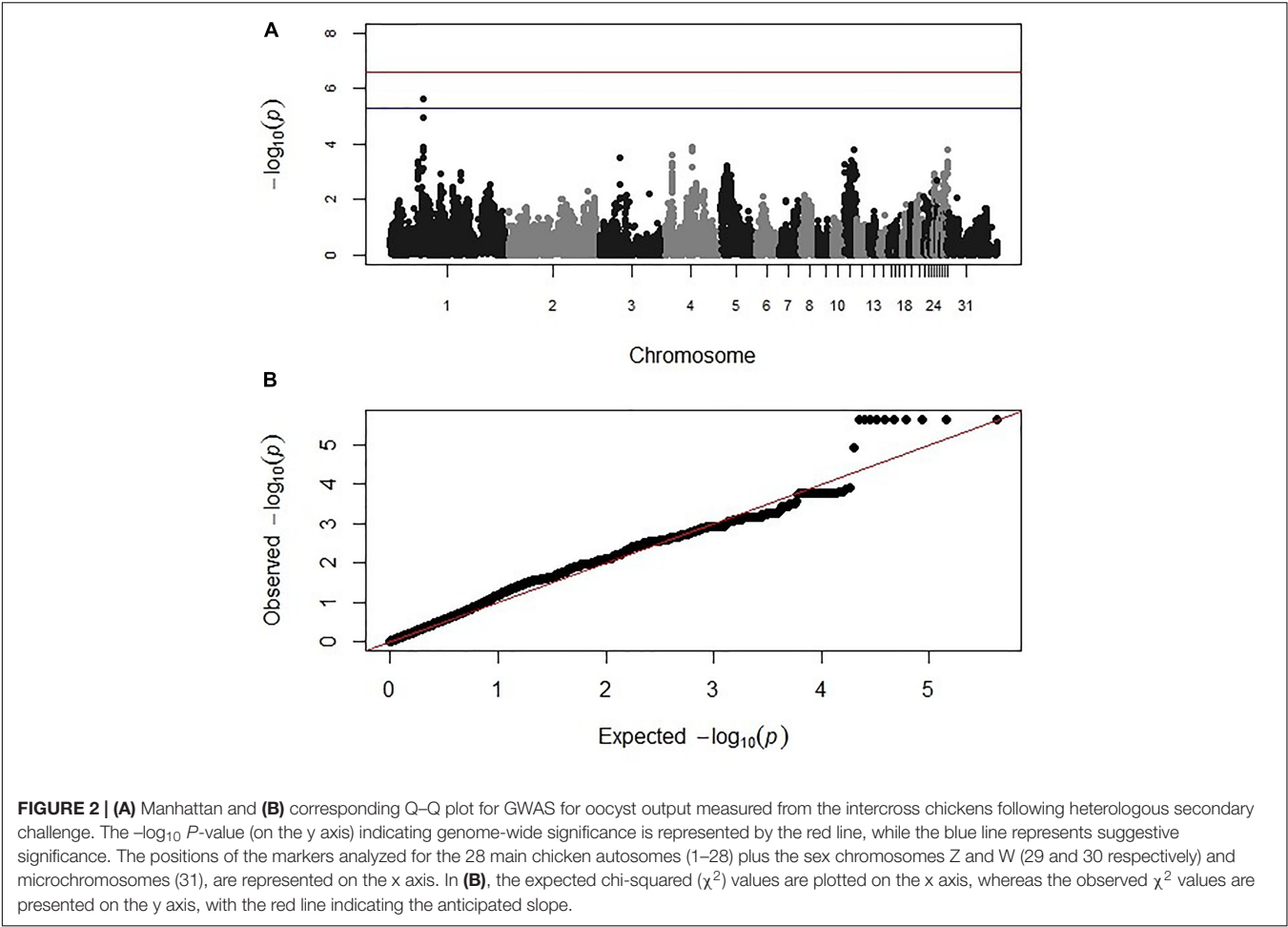
Backcross Study

Genome-wide association study results for resistance to *E. maxima* W primary infection in the backcross population revealed several of significant genomic associations for each of the measured phenotypes. However, there was no overlap of the candidate genomic regions linked to parasite reproduction, intestinal pathology, or IL-10 induction (Figure 3 and Table 3). Specifically, a single marker on chromosome 3 had a suggestive association with PR (Figure 3A and Table 3). Four suggestive marker associations were identified with markers on chromosomes 1, 2, and 3 for intestinal pathology (i.e., lesion

TABLE 1 | Genetic covariance/variance/correlation matrix (± SE), for the backcross primary *E. maxima* infection trial.

| | PR | LS | IL-10 |
|-------|-------------|-------------|-------------|
| PR | 0.92 (0.09) | 0.12 (0.07) | 0.08 (0.07) |
| LS | 0.11 (0.07) | 0.90 (0.09) | 0.15 (0.07) |
| IL-10 | 0.07 (0.07) | 0.14 (0.07) | 1.0 (0.10) |

Significant values are highlighted in italics. Covariances are presented below the diagonal, variances are shaded on the diagonal, with between-trait correlations above. Measured traits are parasite replication per host genome (PR), lesion score (LS), and serum interleukin-10 (IL-10).



damage; **Figure 3B** and **Table 3**). A further four associations were found for IL-10 on chromosomes 1, 2, and 5 (**Figure 3C** and **Table 3**). None of the markers found on chromosome 2 for

TABLE 2 | Details of GWAS-identified and animal model-verified significant markers for oocyst output (OoC) from the intercross chickens following secondary challenge with the heterologous *E. maxima* W strain.

| Trait | Marker | Location (Chr:mb) | G _A (P-value) |
|-------|---------------|-------------------|--------------------------|
| OoC | Affx-50382738 | 1:55145901 | 0.82 (<0.001) |
| | Affx-50382750 | 1:55150313 | 0.82 (<0.001) |
| | Affx-50382770 | 1:55160683 | 0.82 (<0.001) |
| | Affx-50382803 | 1:55177556 | 0.82 (<0.001) |
| | Affx-50382825 | 1:55192552 | 0.82 (<0.001) |
| | Affx-50382867 | 1:55217464 | 0.82 (<0.001) |
| | Affx-50382871 | 1:55220484 | 0.82 (<0.001) |
| | Affx-50382878 | 1:55224572 | 0.82 (<0.001) |
| | Affx-50382880 | 1:55225227 | 0.82 (<0.001) |
| | Affx-50382881 | 1:55225894 | 0.82 (<0.001) |
| | Affx-50382786 | 1:55169757 | 0.82 (<0.001) |
| | | | |
| | | | |

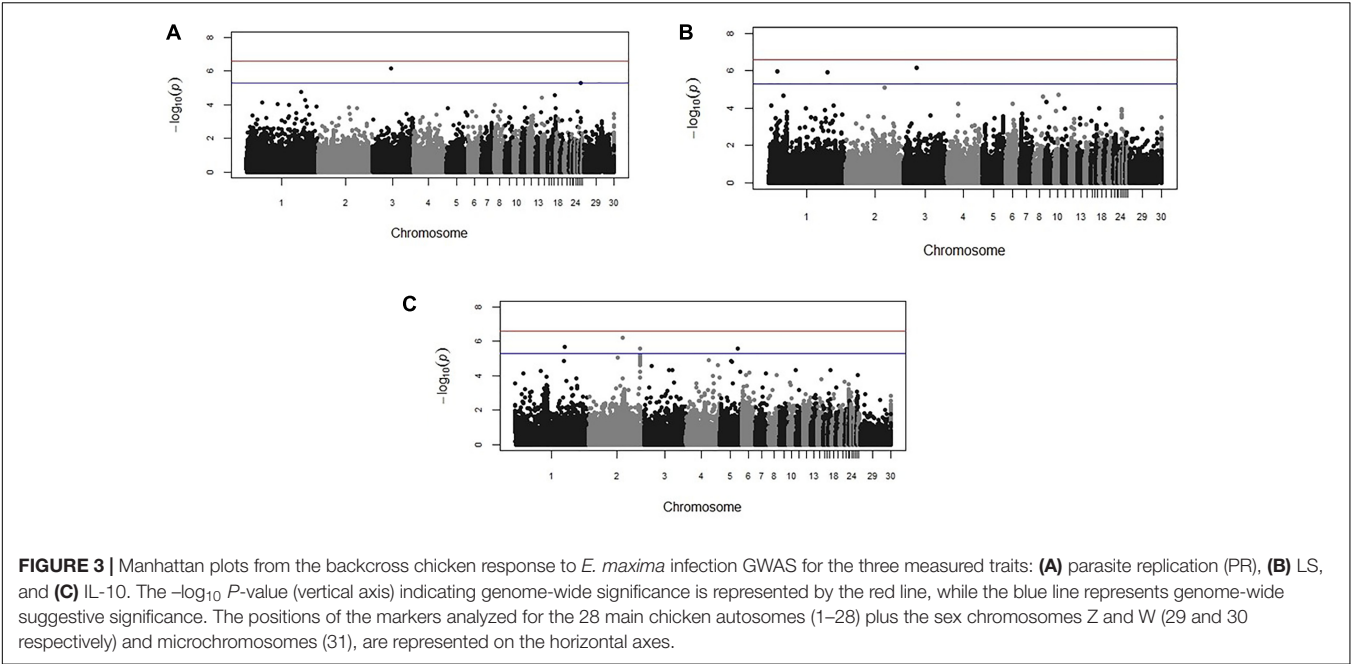
Details provided: Affymetrix marker identifier; chromosome and position of markers in the Galgal5 assembly (Chr:mb); additive genetic effects (G_A), with significance values (P-value).

LS and IL-10 were in common, nor were they in LD. However, the candidate QTL region for IL-10 on chromosome 2 was in proximity with an intercross marker found following primary infection with *E. maxima* H in the intercross population that falls below the suggestive threshold. The corresponding Q-Q plots for GWAS are displayed in **Figure 4**. All significant markers identified in both studies exhibited significant ($P < 0.01$) additive genetic effects (**Table 3**).

All of the significant markers identified for resistance to primary *E. maxima* W infection in the backcross population were located in intronic or intergenic regions (**Supplementary Table S3**). The candidate regions for response to primary *E. maxima* W infection contains a small number of genes: 36 protein-coding genes and four microRNAs (**Supplementary Table S4**).

Resequencing Analysis

In total, 3,230 variants were identified in the candidate regions associated with resistance to primary *E. maxima* infections. SNV located in exonic regions accounted for less than 3% of the total, while the remaining SNV (97%) were located in intronic, upstream, and downstream regions. Genes with SNVs that could potentially lead to non-functional transcripts



were not detected. However, six genes contained missense SNVs that may affect the function of the encoded proteins. More specifically, *LONRF2*, *CHST10*, *PDCL3*, and *TBC1D8* genes on chromosome 1, *FAM69C* on chromosome 2, and *IPCEF1* on chromosome 3 had missense with moderate effect SNVs. Also, these genes contained 3′/5′ UTR variants that may affect the expression of these genes. Details of the missense variants identified in the candidate regions for *E. maxima* resistance to primary infection are presented in **Supplementary Table S5**.

In total, 2,165 SNV were detected in the candidate region on chromosome 1 for the response to heterologous secondary *E. maxima* W challenge. Most of the identified SNV (95%) were located in intronic, upstream and downstream regions; 5% were located in exonic regions, mostly in 3′ and 5′ UTR regions.

TABLE 3 | Details of GWAS-identified and animal model-verified significant markers from the backcross chicken response to *E. maxima* primary infection.

| Trait | Marker | Location Chr:mb | G _A (P-value) |
|-------|---------------|-----------------|--------------------------|
| PR | Affx-51275363 | 3:49284591 | −0.66 (<0.001) |
| LS | Affx-51243371 | 3:31960902 | 0.67 (<0.001) |
| | Affx-50321421 | 1:20945328 | 0.73 (<0.001) |
| | Affx-50226161 | 1:150098597 | 0.57 (<0.001) |
| | Affx-51010702 | 2:98848476 | 0.54 (<0.001) |
| IL-10 | Affx-50999702 | 2:92008902 | −1.49 (<0.001) |
| | Affx-50194384 | 1:132582437 | −1.36 (<0.001) |
| | Affx-51587399 | 5:47737604 | −1.76 (<0.001) |
| | Affx-50857122 | 2:136484786 | 1.42 (0.002) |

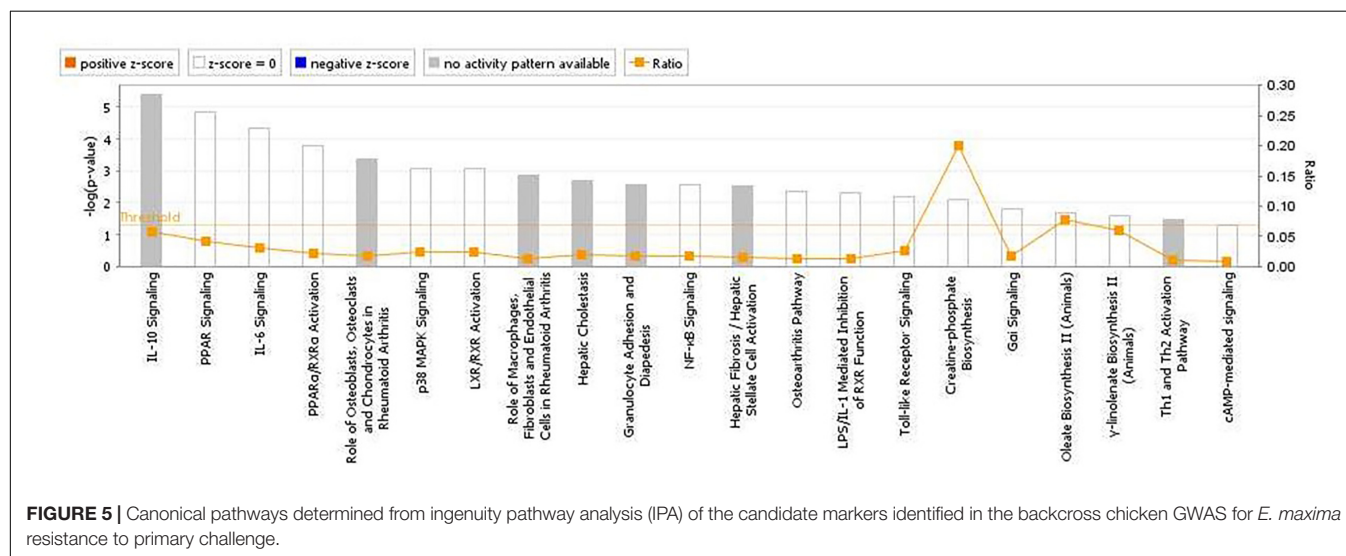
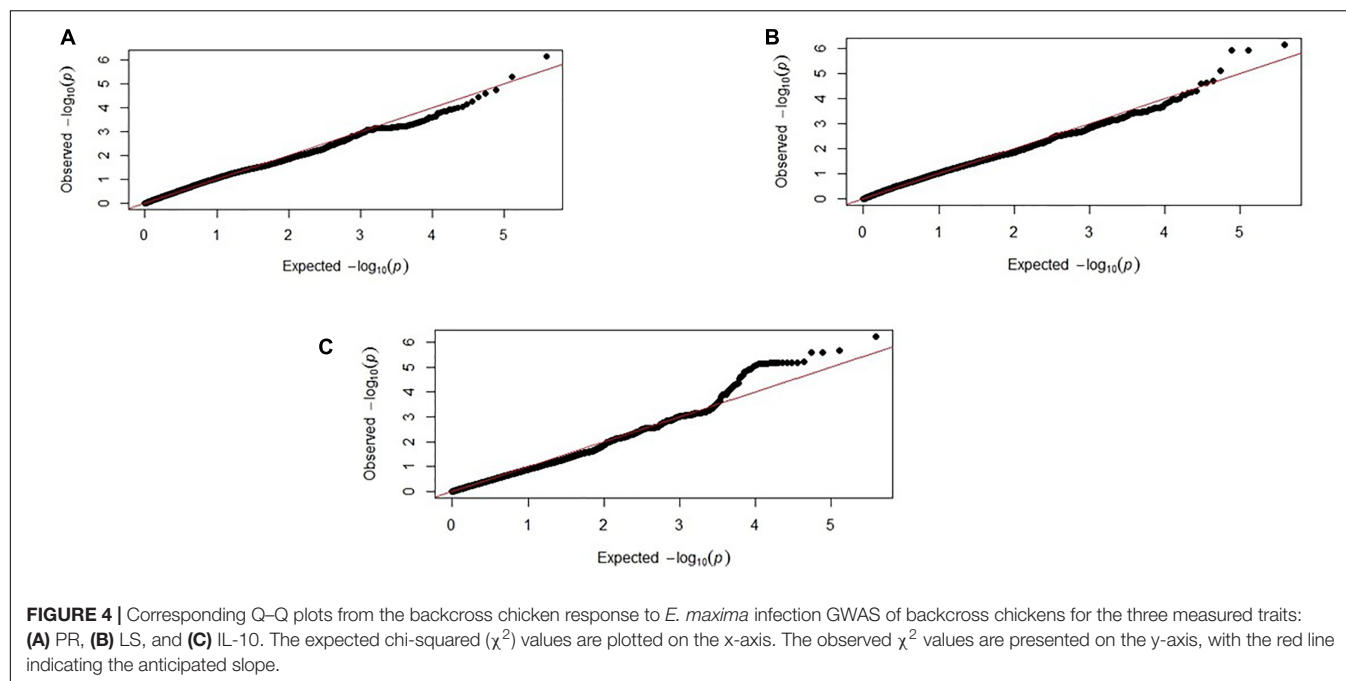
Measured traits – parasite replication per host genome (PR), Lesion Score (LS), and serum interleukin-10 (IL-10). Details provided: Affymetrix marker identifier; chromosome and position of markers in the Gal-gal5 assembly (Chr:mb); the additive genetic effect (G_A) and significance values (P-value).

Nevertheless, three genes (*PMCH*, *TBXAS1*, *THL3*) containing missense variants with moderate effects as well as 3′/5′UTR variants were detected. Details of the missense variants identified in the candidate regions for heterologous secondary *E. maxima* W challenge are presented in **Supplementary Table S6**.

Pathway, Network, and Functional Enrichment Analyses

The analyses for resistance to primary *E. maxima* infection revealed pathway enrichment for immune response involvement, including IL-10, interleukin-6 (IL-6), nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κb) and toll like receptor signaling (**Figure 5**). Using the list of candidate region genes, two networks were constructed, comprising molecular interactions related to inflammatory response and disease, cell death and survival, cellular compromise, and cell cycle (IPA scores = 25; **Figures 6A,B**). A single enriched cluster was found, related to immune response linked to interleukin-1 (IL-1), Toll/IL1 response and cytokine-cytokine receptor response (ES = 2.2, with *IL1R1*, *IL1RL1*, *IL2R*, *IL19R18*, *PTPRM*, and *COL14A* genes involved).

The pathway analyses for response to heterologous *E. maxima* W strain secondary challenge revealed enrichment for both immune (prostanoid biosynthesis, retinoic acid mediated apoptosis signaling, eicosanoid signaling) and metabolic pathways (**Figure 7**). Two gene networks were constructed, related to cell signaling, nucleic acid metabolism and small molecule biochemistry (IPA score = 20), and cellular development, tissue development and function (IPA score = 45), respectively (**Figures 8A,B**). Accompanying functional annotation clustering analysis revealed the presence of two enriched clusters related to cell to cell signaling (ES = 1.7) and metal-ion binding (ES = 1.3).



DISCUSSION

Coccidiosis remains one of the costliest diseases for the international poultry industry. Selectively breeding chickens for enhanced resistance to *Eimeria* challenge, and for improved breadth of vaccine response, could provide a tractable strategy to improve coccidiosis control. We conducted two studies using different crosses between the White Leghorn inbred lines 151 and C.B12. Our data confirm that line 151 birds are more susceptible to primary infection with *E. maxima* than line C.B12 by overall PR (Smith et al., 2002; Blake et al., 2006). While the two inbred lines exhibit similar resistance/susceptibility profiles following primary infection with either of the two antigenically distinct *E. maxima* strains, they show radically

different levels of protection against heterologous secondary challenge by antigenically distinct strains of the same pathogen (Smith et al., 2002). We therefore investigated the genetic background of resistance to primary and heterologous secondary *E. maxima* W challenges.

The resistance of chickens to *Eimeria* infection has traditionally been quantified using measures such as oocyst output and LS, indicating resistance to PR and parasite-induced pathology, respectively. For the former, the fewer oocysts excreted, the more resistant the chicken. Thus, oocyst shedding is considered to be an indicative trait and an accurate phenotype for calculating resistance to primary infection and subsequent parasite challenges and this method was used in the intercross experiment. However, calculation of oocyst

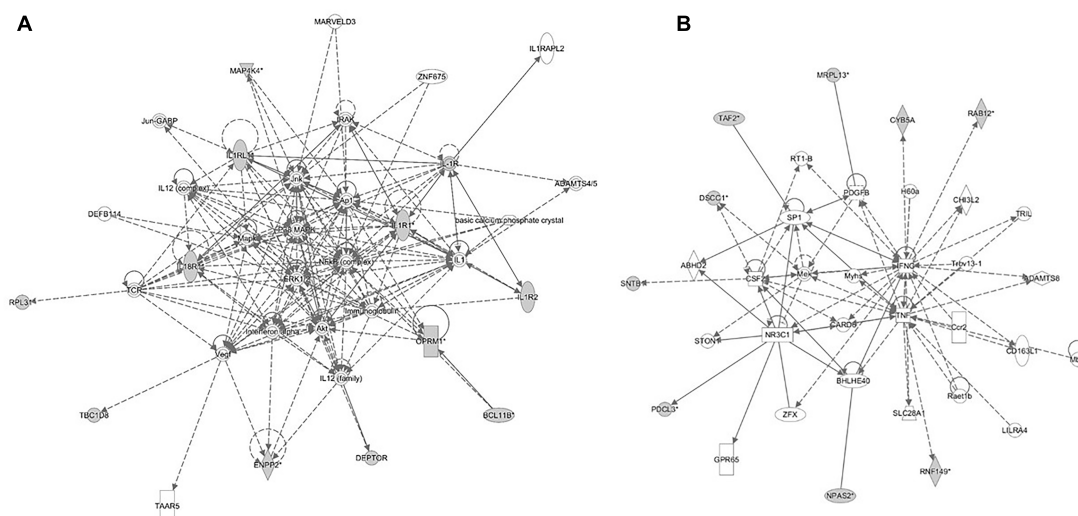


FIGURE 6 | Molecular interaction networks constructed from the canonical pathways identified in the backcross infection response relate to **(A)** inflammatory response and disease and **(B)** cell death and survival, cellular compromise and cell cycle.

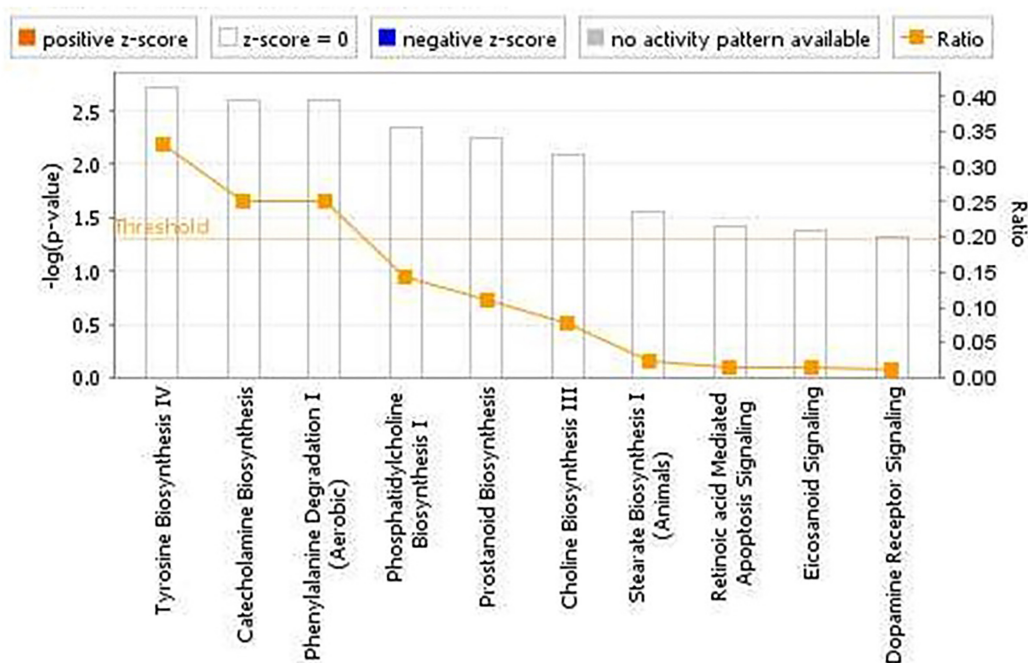
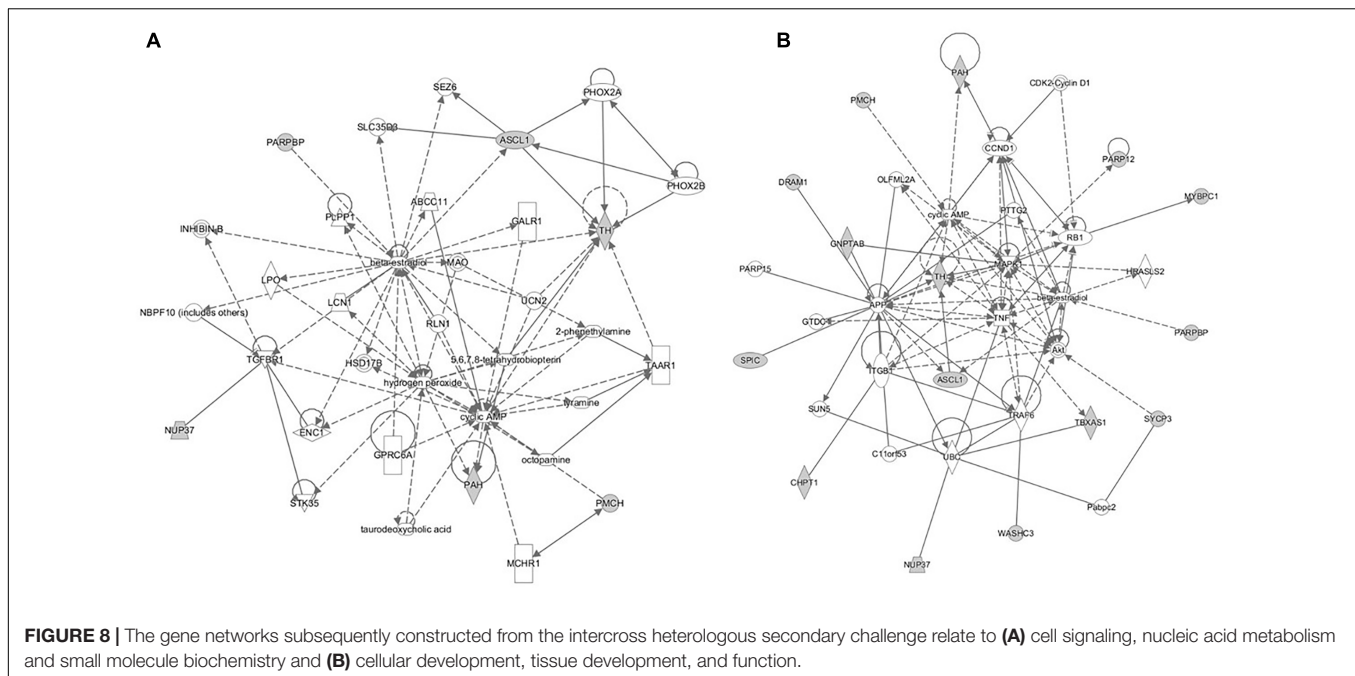


FIGURE 7 | Canonical pathways determined from IPA of the candidate markers identified in the intercross chickens GWAS for *E. maxima* resistance to heterologous secondary challenge.

output by fecal flotation and microscopy is labor intensive. Thus, quantitative real-time PCR for parasite genome copies in intestinal tissues was used as an alternative measure of PR in the more recent backcross experiment (Blake et al., 2006). A third trait, serum IL-10, was also quantified for these latter chickens, providing a measure of the innate immune response to *Eimeria* infection (Rothwell et al., 2004; Boulton et al., 2018). IL-10 is produced after *E. maxima* and *E. tenella*

primary infection of White Leghorn chickens (lines 151 and C.B12) and *E. tenella* primary infection of commercial broilers (Rothwell et al., 2004; Wu et al., 2016; Boulton et al., 2018). In all these cases, IL-10 was expressed at high levels in infected birds only, and significantly correlated with pathology (lesion scores). Here, GWAS from the backcross experiment identified markers associated with IL-10 that exhibit significant additive genetic variance. These findings, in conjunction with



indications that IL-10 is correlated significantly with gross pathology in a commercial population primary infection with *E. tenella* (Boulton et al., 2018), support the use of IL-10 as an accessible early-life biomarker in breeding programs aiming to enhance *Eimeria* resistance to challenge or pathological outcomes.

Although the significance of *E. maxima* in field coccidiosis has been recognized for many years, there has been a limited number of genetic studies investigating host resistance to *E. maxima* primary infection and challenge. A recent study that investigated the genetic background of resistance to high-level *E. maxima* infection using the same HD genotyping array but measuring three different phenotypes (body weight gain, plasma coloration, and β 2-globulin in blood plasma) identified several QTL on chromosomes 1, 2, 3, 5, and 10 in commercial Cobb500 broilers (Hamzic et al., 2015). Similar to our findings, Hamzic et al. (2015) found no QTL overlap among their different phenotypes. Interestingly, QTL identified by Hamzic et al. (2015) on chromosome 1 for β 2-globulin in blood plasma is nearby (2 Mb difference) QTL found in our study linked to for resistance to heterologous secondary *E. maxima* W challenge. Similar enriched biological pathways related to innate immune responses and metabolic processes were also detected in the two studies with this parasite species.

In other comparable work, Zhu et al. (2003) performed a linkage analysis study investigating chicken resistance in terms of oocyst output following controlled *E. maxima* infection using an F2-intercross between two broiler lines with different susceptibility to primary *E. maxima* infection. Using 119 microsatellite markers one locus associated with *E. maxima* resistance was identified on chromosome 1 (Zhu et al., 2003). Expanding this work, Kim et al. (2006) used nine microsatellite markers located on chromosome 1 to refine this region.

According to their results, the peak of QTL was located a considerable genetic distance (i.e., 254 cM) away from the chromosome 1 QTL identified here and in the Hamzic et al. (2015) study. This could be attributed to the use of different chicken lines, *E. maxima* strains, analysis methods, and/or genotyping tools. It is worth mentioning that the power to detect QTL as well as the resolution of their location using a few microsatellites is limited compared to HD genotyping platforms.

Comparison of the re-sequencing data of the two parental chicken lines identified a small number of genes that differ regarding the presence of exonic variants with a putative functional effect on the encoded proteins. Two genes of interest with missense variants located in the candidate regions for resistance to *E. maxima* primary infection encode Phosducin Like 3 (*PDCL3*) and TBC1 Domain Family Member 8 (*TBC1D8*) proteins. These immune-related genes were included in the two networks related to inflammatory response, and cell death and survival, constructed by IPA. *PDCL3* acts as a chaperone for the angiogenic vascular endothelial growth factor receptor, controlling its abundance and inhibiting its ubiquitination and degradation, and also modulating activation of caspases during apoptosis (Wilkinson et al., 2004; Srinivasan et al., 2013). *TBC1D8* is involved in the regulation of cell proliferation, calcium ion transportation, and also has GTPase activator activity (Ishibashi et al., 2009).

The genes encoding Thromboxane A Synthase 1 (*TBXAS1*) and Pro-Melanin Concentrating Hormone (*PMCH*) are located in the candidate region and are of interest in resistance to secondary challenge by heterologous *E. maxima* W. *TBXAS1* encodes a member of the cytochrome P450 superfamily of enzymes involved in both immune response and metabolism; it plays a role in drug metabolism, platelet activation and

metabolism, and synthesis of cholesterol, steroids, and other lipids (Yokoyama et al., 1991; Miyata et al., 1994). The pro-inflammatory actions of thromboxane receptors have been demonstrated to enhance cellular immune responses in a mouse model (Thomas et al., 2003). *PMCH* encodes a preproprotein that is proteolytically processed to generate multiple protein products, including melanin-concentrating hormone (MCH) that stimulates hunger and may additionally regulate energy homeostasis, reproductive function, and sleep (Viale et al., 1997; Chagnon et al., 2007). In a further mouse model, MCH has also been reported as a mediator of intestinal inflammation (Kokkotou et al., 2008). Although, the genes mentioned above are good functional candidates for resistance to primary infection and heterologous challenge with *E. maxima*, further studies are needed to confirm the present results and identify the actual causative genes and mutations.

The immune interactions between an intracellular pathogen and a host are complex and vary as a consequence of the survival mechanisms that have evolved in both (Blake et al., 2011; Blake and Tomley, 2014). It has been suggested that host control of challenge with *Eimeria*, an obligate intracellular pathogen, requires a strong inflammatory, mostly cell mediated response (Shirley et al., 2005; Dalloul and Lillehoj, 2006). Also, host innate immune responses have been detected during initial pathogen exposure in several studies (Kim et al., 2008; Pinard-van der Laan et al., 2009; Wu et al., 2016; Boulton et al., 2018). According to our findings, several gene networks and pathways relating to innate, humoral and cell-mediated, immune responses were highlighted from the gene products located in the candidate regions for resistance to primary *Eimeria* infection. Among the canonical pathways, IL-10 signaling was the most significant, with relevance as a regulator of cytokines such as interferon- (IFN-) γ . These findings agree with previous studies of *Eimeria* resistance that have highlighted $\text{IFN } \gamma$ and tumor necrosis factor (TNF) nodes as crucial (Pinard-Van Der Laan et al., 1998; Smith and Hayday, 2000a,b; Bacciu et al., 2014), since IL-10 downregulates $\text{IFN}\gamma$ production (Schaefer et al., 2009).

CONCLUSION

We identified genomic regions, putative candidate genes, canonical pathways and networks involved in the underlying molecular mechanisms of chicken resistance to *E. maxima* primary infection and to secondary heterologous *E. maxima* strain challenge. More emphasis should be placed on the relevant mechanisms for disease resistance, the response to secondary heterologous strain challenge and the role of IL-10 induction in immune responses to intestinal challenge in the future selective breeding of chickens.

REFERENCES

- Aulchenko, Y. S., Ripke, S., Isaacs, A., and Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi: 10.1093/bioinformatics/btm108
- Bacciu, N., Bed'hom, B., Filangi, O., Rome, H., Gourichon, D., Reperant, J. M., et al. (2014). QTL detection for coccidiosis (*Eimeria tenella*) resistance in a fayoumi x leghorn F_2 cross, using a medium-density SNP panel. *Genet. Sel. Evol.* 46:14. doi: 10.1186/1297-9686-46-14
- Blake, B. J., and Tomley, A. M. (2014). The immune response to *Eimeria* infections. *Parasitology* 144, 1–14. doi: 10.1017/S0022278X13000888
- Blake, B. J., Tomley, A. M., and Nisbet, A. M. (2011). The immune response to *Eimeria* infections. *Parasitology* 141, 1–14. doi: 10.1017/S0022278X10000888
- Chagnon, M., and Vigne, R. (2007). The melanin-concentrating hormone (MCH) system: a key player in the regulation of energy homeostasis, reproductive function, and sleep. *Neurosci. Biobehav. Rev.* 31, 1–14. doi: 10.1016/j.neubiorev.2006.08.001
- Kim, H. S., Pinard-van der Laan, A., and Wu, Y. (2008). The immune response to *Eimeria* infections. *Parasitology* 138, 1–14. doi: 10.1017/S0022278X08000888
- Kokkotou, S., and Karamaouna, M. (2008). The immune response to *Eimeria* infections. *Parasitology* 138, 1–14. doi: 10.1017/S0022278X08000888
- Pinard-Van Der Laan, A., Smith, J. E., and Hayday, A. C. (1998). The immune response to *Eimeria* infections. *Parasitology* 126, 1–14. doi: 10.1017/S0022278X98000888
- Pinard-Van Der Laan, A., Smith, J. E., and Hayday, A. C. (2000a,b). The immune response to *Eimeria* infections. *Parasitology* 120, 1–14. doi: 10.1017/S0022278X00000888
- Pinard-Van Der Laan, A., Smith, J. E., and Hayday, A. C. (2014). The immune response to *Eimeria* infections. *Parasitology* 144, 1–14. doi: 10.1017/S0022278X13000888
- Shirley, B. J., and Tomley, A. M. (2005). The immune response to *Eimeria* infections. *Parasitology* 130, 1–14. doi: 10.1017/S0022278X05000888
- Smith, J. E., and Hayday, A. C. (2000a,b). The immune response to *Eimeria* infections. *Parasitology* 120, 1–14. doi: 10.1017/S0022278X00000888
- Smith, J. E., and Hayday, A. C. (2014). The immune response to *Eimeria* infections. *Parasitology* 144, 1–14. doi: 10.1017/S0022278X13000888
- Thomas, J. L., and Hayday, A. C. (2003). The immune response to *Eimeria* infections. *Parasitology* 126, 1–14. doi: 10.1017/S0022278X03000888
- Viale, R. M., and Vigne, R. (1997). The melanin-concentrating hormone (MCH) system: a key player in the regulation of energy homeostasis, reproductive function, and sleep. *Neurosci. Biobehav. Rev.* 21, 1–14. doi: 10.1016/S0304-3940(97)00088-8
- Wu, Y., Pinard-van der Laan, A., and Kim, H. S. (2016). The immune response to *Eimeria* infections. *Parasitology* 146, 1–14. doi: 10.1017/S0022278X15000888
- Yokoyama, C., and Miyata, T. (1991). The melanin-concentrating hormone (MCH) system: a key player in the regulation of energy homeostasis, reproductive function, and sleep. *Neurosci. Biobehav. Rev.* 15, 1–14. doi: 10.1016/0304-3940(91)90088-8

AVAILABILITY OF SUPPORTING DATA

The resequencing data used in this study is available in NCBI dbSNP at the following web page: http://www.ncbi.nlm.nih.gov/SNP/snp_viewBatch.cgi?sbid=1062063.

AUTHOR CONTRIBUTIONS

AS, PK, SB, FT, and DB devised the overall strategy and obtained funding. PK, SB, FT, and DB conceived the backcross experiments. PMH and KB devised the backcross breeding. MN managed the backcross trials and performed qPCR and DNA extraction assisted by KH and KB. Backcross phenotype collection was carried out by MN, KH and KB, while DB scored lesions. ZW performed IL-10 assays assisted by KB. KB prepared backcross DNA for genotyping and carried out all backcross analyses with input from AP, VR, and OM. AS designed the intercross trials with input from NB and these were carried out by PH and AA. AP performed an initial analysis of the intercross data with input from OM and KB. Pathway and resequencing analyses were performed by AP and KB. The manuscript was drafted by KB and AP with input from all other authors except PMH, SB, NB, and PK. AS, DB, FT, DH, and AP assisted in the interpretation of results.

FUNDING

The backcross work was funded by the BBSRC through the Animal Research Club (ARC) program under grants BB/L004046 and BB/L004003, while DEFRA OD0534 and BBSRC BB/E01089X/1 funded the intercross study.

ACKNOWLEDGMENTS

We thank our colleagues who assisted in the collection of phenotypes, including Iván Pastor-Fernández, Lucy Freem, Angela Stebbings and Nigel Salmon, and the staff at the NARF Bumstead, RVC animal welfare barn and Institute for Animal Health (now Pirbright Institute), Compton facilities. We also thank Richard Kuo for providing the re-sequencing information of the two inbred chicken lines.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00528/full#supplementary-material>

- Blake, D. P., Billington, K. J., Copestake, S. L., Oakes, R. D., Quail, M. A., Wan, K. L., et al. (2011). Genetic mapping identifies novel highly protective antigens for an apicomplexan parasite. *PLoS Pathog.* 7:e1001279. doi: 10.1371/journal.ppat.1001279
- Blake, D. P., Hesketh, P., Archer, A., Carroll, F., Shirley, M. W., and Smith, A. L. (2005). The influence of immunizing dose size and schedule on immunity to subsequent challenge with antigenically distinct strains of *Eimeria maxima*. *Avian Pathol.* 34, 489–494. doi: 10.1080/03079450500368292
- Blake, D. P., Hesketh, P., Archer, A., Carroll, F., Smith, A. L., and Shirley, M. W. (2004). Parasite genetics and the immune host: recombination between antigenic types of *Eimeria maxima* as an entrée to the identification of protective antigens. *Mol. Biochem. Parasitol.* 138, 143–152. doi: 10.1016/j.molbiopara.2004.08.006
- Blake, D. P., Hesketh, P., Archer, A., Shirley, M. W., and Smith, A. L. (2006). *Eimeria maxima*: the influence of host genotype on parasite reproduction as revealed by quantitative real-time PCR. *Int. J. Parasitol.* 36, 97–105. doi: 10.1016/j.ijpara.2005.09.011
- Blake, D. P., and Tomley, F. M. (2014). Securing poultry production from the ever-present *Eimeria* challenge. *Trends Parasitol.* 30, 12–19. doi: 10.1016/j.pt.2013.10.003
- Boulton, K., Nolan, M. J., Wu, Z., Psifidi, A., Harman, K., Bishop, S. C., et al. (2018). Phenotypic and genetic variation in the response of chickens to *Eimeria tenella* induced coccidiosis. *Genet. Sel. Evol.* 50:63. doi: 10.1186/s12711-018-0433-7
- Bumstead, N., and Millard, B. (1987). Genetics of resistance to coccidiosis: response of inbred chicken lines to infection by *Eimeria tenella* and *Eimeria maxima*. *Br. Poult. Sci.* 28, 705–715. doi: 10.1080/00071668708417006
- Bumstead, N., and Millard, B. J. (1992). Variation in susceptibility of inbred lines of chickens to 7 species of *Eimeria*. *Parasitology* 104, 407–413. doi: 10.1017/S0031182000063654
- Chagnon, Y. C., Bureau, A., Gendron, D., Bouchard, R. H., Merette, C., Roy, M. A., et al. (2007). Possible association of the pro-melanin-concentrating hormone gene with a greater body mass index as a side effect of the antipsychotic olanzapine. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 144B, 1063–1069. doi: 10.1002/ajmg.b.30554
- Crouch, C. F., Andrews, S. J., Ward, R. G., and Francis, M. J. (2003). Protective efficacy of a live attenuated anti-coccidial vaccine administered to 1-day-old chickens. *Avian Pathol.* 32, 297–304. doi: 10.1080/10307945031000097912
- Dalloul, R. A., and Lillehoj, H. S. (2006). Poultry coccidiosis: recent advancements in control measures and vaccine development. *Expert Rev. Vaccines* 5, 143–163. doi: 10.1586/14760584.5.1.143
- Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:P3. doi: 10.1186/gb-2003-4-5-p3
- Eckert, J., Brown, R., Shirley, M. W., and Coudert, P. (1995). *Guidelines on Techniques in Coccidiosis Research*. Brussels: European Commission.
- Fitz-Coy, S. (1992). Antigenic variation among strains of *Eimeria maxima* and *E. tenella* of the Chicken. *Avian Dis.* 36, 40–43. doi: 10.2307/1591712
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., and Thompson, R. (2015). *ASReml User Guide Release 4.1 Structural Specification*. Hemel Hempstead: VSN International Ltd.
- Hamzic, E., Buitenhuis, B., Herault, F., Hawken, R., Abrahamsen, M. S., Servin, B., et al. (2015). Genome-wide association study and biological pathway analysis of the *Eimeria maxima* response in broilers. *GSE* 47:91. doi: 10.1186/s12711-015-0170-0
- Hong, Y. H., Lillehoj, H. S., Lillehoj, E. P., and Lee, S. H. (2006). Changes in immune-related gene expression and intestinal lymphocyte subpopulations following *Eimeria maxima* infection of chickens. *Vet. Immunol. Immunopathol.* 114, 259–272. doi: 10.1016/j.vetimm.2006.08.006
- Ishibashi, K., Kanno, E., Itoh, T., and Fukuda, M. (2009). Identification and characterization of a novel Tre-2/Bub2/Cdc16 (TBC) protein that possesses Rab3A-GAP activity. *Genes Cells* 14, 41–52. doi: 10.1111/j.1365-2443.2008.01251.x
- Johnson, C. A., Kennedy, T. J., and Moeller, M. W. (1986). “Immunization of chickens against coccidiosis by termination of infections with Bay VI 9142,” in *Proceedings of the Georgia Coccidiosis Conference, Research in Avian Coccidiosis* (Athens: University of Georgia), 253–262.
- Johnson, J., and Reid, W. M. (1970). Anticoccidial drugs: lesion scoring techniques in battery and floor-pen experiments with chickens. *Exp. Parasitol.* 28, 30–36. doi: 10.1016/0014-4894(70)90063-9
- Kim, D. K., Lillehoj, H. S., Hong, Y. H., Park, D. W., Lamont, S. J., Han, J. Y., et al. (2008). Immune-related gene expression in two B-complex disparate genetically inbred fayoumi chicken lines following *eimeria maxima* infection. *Poult. Sci.* 87, 433–443. doi: 10.3382/ps.2007-00383
- Kim, E. S., Hong, Y. H., Min, W., and Lillehoj, H. S. (2006). Fine-mapping of coccidia-resistant quantitative trait loci in chickens. *Poult. Sci.* 85, 2028–2030. doi: 10.1093/ps/85.11.2028
- Kokkotou, E., Moss, A. C., Torres, D., Karagiannides, I., Cheifetz, A., Liu, S., et al. (2008). Melanin-concentrating hormone as a mediator of intestinal inflammation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10613–10618. doi: 10.1073/pnas.0804536105
- Krämer, A., Green, J., Pollard, J. J., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703
- Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14:59. doi: 10.1186/1471-2164-14-59
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- McDonald, V., and Shirley, M. W. (2009). Past and future: vaccination against *Eimeria*. *Parasitology* 136, 1477–1489. doi: 10.1017/S0031182009006349
- Miyata, A., Yokoyama, C., Ihara, H., Bandoh, S., Takeda, O., Takahashi, E., et al. (1994). Characterization of the human gene (TBXAS1) encoding thromboxane synthase. *Eur. J. Biochem.* 224, 273–279. doi: 10.1111/j.1432-1033.1994.00273.x
- Nolan, M. J., Tomley, F. M., Kaiser, P., and Blake, D. P. (2015). Quantitative real-time PCR (qPCR) for *Eimeria tenella* replication — implications for experimental refinement and animal welfare. *Parasitol. Int.* 64, 464–470. doi: 10.1016/j.parint.2015.06.010
- Norton, C. C., and Hein, H. E. (1976). *Eimeria maxima*: a comparison of two laboratory strains with a fresh isolate. *Parasitology* 72, 345–354. doi: 10.1017/S0031182000049544
- Pinard-van der Laan, M. H., Bed'hom, B., Coville, J.-L., Pitel, F., Feve, K., Leroux, S., et al. (2009). Microsatellite mapping of QTLs affecting resistance to coccidiosis (*Eimeria tenella*) in a fayoumi x white leghorn cross. *BMC Genomics* 10:31. doi: 10.1186/1471-2164-10-31
- Pinard-Van Der Laan, M. H., Monvoisin, J. L., Pery, P., Hamet, N., and Thomas, M. (1998). Comparison of outbred lines of chickens for resistance to experimental infection with coccidiosis (*Eimeria tenella*). *Poult. Sci.* 77, 185–191. doi: 10.1093/ps/77.2.185
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Reid, A. J., Blake, D. P., Ansari, H. R., Billington, K., Browne, H. P., Bryant, J., et al. (2014). Genomic analysis of the causative agents of coccidiosis in domestic chickens. *Genome Res.* 24, 1676–1685. doi: 10.1101/gr.168955.113
- Rothwell, L., Young, J. R., Zoorob, R., Whittaker, C. A., Hesketh, P., Archer, A., et al. (2004). Cloning and characterization of chicken IL-10 and its role in the immune response to *Eimeria maxima*. *J. Immunol.* 173, 2675–2682. doi: 10.4049/jimmunol.173.4.2675
- Schaefer, A., Unterberger, C., Frankenberger, M., Lohrum, M., Staples, K. J., Werner, T., et al. (2009). Mechanism of interferon-gamma mediated down-regulation of interleukin-10 gene expression. *Mol. Immunol.* 46, 1351–1359. doi: 10.1016/j.molimm.2008.11.015
- Shirley, M. W., Smith, A. L., and Tomley, F. M. (2005). “The biology of avian *Eimeria* with an emphasis on their control by vaccination,” in *Advances in Parasitology*, Vol. 60, eds J. R. Baker, R. Muller, and D. Rollinson (Amsterdam: Elsevier Ltd), 285–330.
- Smith, A. L., and Hayday, A. C. (2000b). An $\alpha\beta$ T-cell-independent immunoprotective response towards gut coccidia is supported by $\gamma\delta$ cells. *Immunology* 101, 325–332.
- Smith, A. L., and Hayday, A. C. (2000a). Genetic dissection of primary and secondary responses to a widespread natural pathogen of the gut, *Eimeria veriformis*. *Infect. Immun.* 68, 6273–6280.

- Smith, A. L., Hesketh, P., Archer, A., and Shirley, M. W. (2002). Antigenic diversity in *Eimeria maxima* and the influence of host genetics and immunization schedule on cross-protective immunity. *Infect. Immun.* 70, 2472–2479. doi: 10.1128/IAI.70.5.2472-2479.2002
- Srinivasan, S., Meyer, R. D., Lugo, R., and Rahimi, N. (2013). Identification of PDCL3 as a novel chaperone protein involved in the generation of functional VEGF receptor 2. *J. Biol. Chem.* 288, 23171–23181. doi: 10.1074/jbc.M113.473173
- Thomas, D. W., Rocha, P. N., Nataraj, C., Robinson, L. A., Spurney, R. F., Koller, B. H., et al. (2003). Proinflammatory actions of thromboxane receptors to enhance cellular immune responses. *J. Immunol.* 171, 6389–6395. doi: 10.4049/jimmunol.171.12.6389
- Viale, A., Zhixing, Y., Breton, C., Pedetour, F., Coquerel, A., Jordan, D., et al. (1997). The melanin-concentrating hormone gene in human: flanking region analysis, fine chromosome mapping, and tissue-specific expression. *Brain Res. Mol. Brain Res.* 46, 243–255. doi: 10.1016/S0169-328X(97)00018-1
- Wilkinson, J. C., Richter, B. W., Wilkinson, A. S., Burstein, E., Rumble, J. M., Balliu, B., et al. (2004). VIAF, a conserved inhibitor of apoptosis (IAP)-interacting factor that modulates caspase activation. *J. Biol. Chem.* 279, 51091–51099. doi: 10.1074/jbc.M409623200
- Wu, Z., Hu, T., Rothwell, L., Vervelde, L., Boulton, K., Nolan, M. J., et al. (2016). Analysis of the function of IL-10 in chickens using specific neutralising antibodies and a sensitive capture ELISA. *Dev. Comp. Immunol.* 63, 206–212. doi: 10.1016/j.dci.2016.04.016
- Yokoyama, C., Miyata, A., Ihara, H., Ullrich, V., and Tanabe, T. (1991). Molecular cloning of human platelet thromboxane A synthase. *Biochem. Biophys. Res. Commun.* 178, 1479–1484. doi: 10.1016/0006-291X(91)91060-P
- Zhu, J., Lillehoj, H. S., Allen, P. C., Van Tassell, C. P., Sonstegard, T. S., Cheng, H. H., et al. (2003). Mapping quantitative trait loci associated with resistance to coccidiosis and growth. *Poult. Sci.* 82, 9–16. doi: 10.1093/ps/82.1.9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with several of the authors, NB, PH, AA, and AS at time of review.

Copyright © 2018 Boulton, Nolan, Wu, Riggio, Matika, Harman, Hocking, Bumstead, Hesketh, Archer, Bishop, Kaiser, Tomley, Hume, Smith, Blake and Psifidi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Novel Resilience Phenotypes Using Feed Intake Data From a Natural Disease Challenge Model in Wean-to-Finish Pigs

Austin M. Putz¹, John C. S. Harding², Michael K. Dyck³, F. Fortin⁴, Graham S. Plastow³, Jack C. M. Dekkers^{1*} and PigGen Canada[†]

¹ Department of Animal Science, Iowa State University, Ames, IA, United States, ² Department of Large Animal Clinical Sciences, University of Saskatchewan, Saskatoon, SK, Canada, ³ Department of Agriculture, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, ⁴ Centre de Développement du Porc du Québec Inc. (CDPQ), Québec City, QC, Canada

OPEN ACCESS

Edited by:

Andrea B. Doeschl-Wilson,
University of Edinburgh,
United Kingdom

Reviewed by:

Ilias Kyriazakis,
Newcastle University, United Kingdom
Allan Schinckel,
Purdue University, United States
Susanne Hermes,
University of New England, Australia

*Correspondence:

Jack C. M. Dekkers
jdekke@iastate.edu

[†] PigGen Canada authors are listed at
the end of the article

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 24 April 2018

Accepted: 04 December 2018

Published: 08 January 2019

Citation:

Putz AM, Harding JCS, Dyck MK,
Fortin F, Plastow GS, Dekkers JCM
and PigGen Canada (2019) Novel
Resilience Phenotypes Using Feed
Intake Data From a Natural Disease
Challenge Model in Wean-to-Finish
Pigs. *Front. Genet.* 9:660.
doi: 10.3389/fgene.2018.00660

The objective of this study was to extract novel phenotypes related to disease resilience using daily feed intake data from growing pigs under a multifactorial natural disease challenge that was designed to mimic a commercial environment with high disease pressure to maximize expression of resilience. Data used were the first 1,341 crossbred wean-to-finish pigs from a research facility in Québec, Canada. The natural challenge was established under careful veterinary oversight by seeding the facility with diseased pigs from local health-challenged farms, targeting various viral and bacterial diseases, and maintaining disease pressure by entering batches of 60–75 pigs in a continuous flow system. Feed intake (FI) is sensitive to disease, as pigs tend to eat less when they become ill. Four phenotypes were extracted from the individual daily FI data during finishing as novel measures of resilience. The first two were daily variability in FI or FI duration, quantified by the root mean square error (RMSE) from the within individual regressions of FI or duration at the feeder (DUR) on age (RMSE_{FI} and RMSE_{DUR}). The other two were the proportion of off-feed days, classified based on negative residuals from a 5% quantile regression (QR) of daily feed intake or duration data on age across all pigs (QR_{FI} and QR_{DUR}). Mortality and treatment rate had a heritability of 0.13 (± 0.05) and 0.29 (± 0.07), respectively. Heritability estimates for RMSE_{FI}, RMSE_{DUR}, QR_{FI}, and QR_{DUR} were 0.21 (± 0.07), 0.26 (± 0.07), 0.15 (± 0.06), and 0.23 (± 0.07), respectively. Genetic correlations of RMSE and QR measures with mortality and treatment rate ranged from 0.37 to 0.85, with QR measures having stronger correlations with both. Estimates of genetic correlations of RMSE measures with production traits were typically low, but often favorable (e.g., -0.31 between RMSE_{FI} and finishing ADG). Although disease resilience was our target, fluctuations in FI and duration can be caused by many factors other than disease and should be viewed as overall indicators of general resilience to a variety of stressors. In conclusion, daily variation in FI or duration at the feeder can be used as heritable measures of resilience.

Keywords: resilience, disease resistance, feed intake, feeding duration, pigs

INTRODUCTION

Disease resilience can be defined as the ability to maintain relatively undiminished performance levels under infection (Albers et al., 1987; Doeschl-Wilson et al., 2012; Mulder and Rashidi, 2017). In the literature, much focus has been placed on separating disease resistance and tolerance (Bishop, 2012; Bishop and Woolliams, 2014; Lough et al., 2015). Disease resilience is an alternative to selection for a combination of resistance and tolerance (Guy et al., 2012; Mulder and Rashidi, 2017). Most studies on resilience (e.g., Mulder and Rashidi, 2017), however, consider only a single disease but an animal could be resistant or tolerant to one disease and more susceptible to other diseases. Currently, there are dozens of pathogens for swine around the world, including viral, bacterial, and parasitic infectious diseases (Zimmerman et al., 2012). Pathogens can be spread around the world. New pathogens and alternative strains will continue to develop as well. Breeding companies that market breeding stock across the globe have to simultaneously consider disease resilience to many of these pathogens and environments. Selecting animals that maintain performance in a typical commercial system provides a natural weighting of resilience to each disease based on the impact of each disease on productivity, along with the incidence or prevalence of the disease. van der Waaij et al. (2000) stated that observed production can be viewed as a selection index where the underlying components are weighted based on their impacts on performance. It is important, however, that the testing environment is representative of the target commercial environments. Resilience can be an effective, but “black-box” approach to selection for disease resistance and tolerance in animals (Mulder and Rashidi, 2017). One of the challenges, however, is to obtain heritable measures or indicators of resilience for selection, as elite breeding populations are typically kept in high-health conditions.

Recently, Elgersma et al. (2018) exploited routinely collected daily milk yield to quantify resilience in lactating dairy cows because daily milk yield is sensitive to diseases such as mastitis. Both significant drops in milk yield and day-to-day variation in milk yield within cow were used to quantify resilience. These phenotypes did not quantify disease resilience specifically, as it was not possible to validate that all changes in milk yield were related to infectious diseases. This becomes a multifactorial issue as causes for drops in milk yield can include mastitis, lameness, subclinical ketosis, and displaced abomasum, among others (King et al., 2018). This leads to these types of phenotypes capturing disease resilience along with general resilience (Elgersma et al., 2018). When selection for growth under a high stress environment was practiced in cattle, Frisch (1981) found that the selected animals were more productive under challenge but that this selection did not change their growth potential. If the goal is to only target disease resilience, this is a disadvantage for measuring production or deviations in production. For instance, in dairy cattle, using somatic cell count as an indicator trait may be better for selection against only mastitis than measuring productivity fluctuations in milk yield or feed intake. However, if the breeding objective is to maintain productivity regardless of the causes associated with

milk yield deviations (i.e., general resilience), phenotypes that measure changes in productivity over time within animal are likely to have an economic value themselves (Elgersma et al., 2018).

Much is known about the relationship between feed intake (FI) and anorexia (Sandberg et al., 2006; Kyriazakis and Doeschl-Wilson, 2009). Production of cytokines such as interleukin-6 (IL-6) and tumor necrosis factor- α (TNF- α) can cause a loss of appetite (Webel et al., 1997; Petry et al., 2007; Kyriazakis and Doeschl-Wilson, 2009). Knap (2009) suggested that individual day-to-day variation in feed intake could be utilized to quantify environmental sensitivity such as resilience to heat stress. Animals with more day-to-day variation in FI would indicate animals that are less resilient. Under a disease challenge, day-to-day variation in FI would reflect resilience to disease.

Alternative feed intake traits from individual FI electronic systems have been analyzed previously for the purpose of developing indicator traits for feed intake or feed efficiency in a selection index (de Haer et al., 1993; Von Felde et al., 1996; Schulze et al., 2003; Young et al., 2011; Lu et al., 2017). The most common and simplest of these traits investigated are occupation time at the feeder (or duration), number of visits, and FI rate (kg feed / unit time). Other feeding traits during the course of a day have also been investigated (Kyriazakis and Tolcamp, 2018). Individual FI is typically recorded in high-health environments, which limits the use of these data in nucleus herds to quantify traits related to environmental sensitivity or resilience (mostly due to health). FI traits such as feeding duration (i.e., time at the feeder) could also exhibit day-to-day variability from causes such as illness and may be a more feasible alternative to collecting individual FI in these challenged environments if typical commercial feeders could be enhanced with antennae to collect time at the feeder on individual pigs (with RFID tags). Feeding traits, such as duration, become more valuable in severely challenged environments due to the fact that if a pig stops eating completely their time at the feeder is expected to be zero.

The objectives of this study were to (1) develop and evaluate novel measures of resilience based on daily feed intake and feeding duration data for finishing pigs in a health-challenged environment and (2) determine heritabilities and genetic correlations of these measures with mortality, treatments, and other economically important production traits.

MATERIALS AND METHODS

This study was carried out in accordance with the recommendations of the Canadian Council on Animal Care (<https://www.cac.ca/en/certification/about-certification/>). The protocol was approved by the Protection Committee of the Centre de Recherche en Sciences Animales de Deschambault (CRSAD; <http://www.crsad.qc.ca/>). The Centre de développement du porc du Québec (CDPQ) had full oversight on the project along with veterinarians.

Natural Challenge Protocol

A natural challenge wean-to-finish protocol was established in late 2015 at CDPQ in Québec, Canada, with the aim to mimic

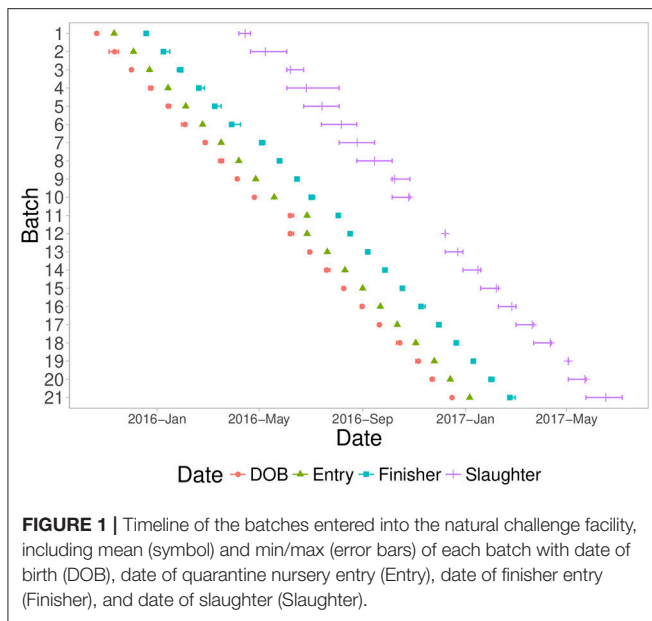
a commercial farm with high disease pressure to maximize expression of genetic differences in resilience. The protocol was established at a research facility to allow detailed phenotype recording, blood sampling, and *in vivo* assays. This is an ongoing project that will conclude in early 2019. The natural challenge facility consists of three consecutive phases: (1) a healthy quarantine nursery for ~19 days after weaning, (2) a late nursery phase, where pigs are first exposed to disease for ~4 weeks, and (3) a finishing phase for the remainder of the growing period (69–181 days of age on average). Phases 2 and 3 are in the same barn, connected by a hallway and are collectively referred to as the “challenge facility.” Phase 1 is at a nursery approximately 1 km south of the challenge facility and is kept free of disease using strict biosecurity between the facilities. In the quarantine nursery, samples, and measurements are taken for future development of early predictors of resilience in a non-challenged environment, typical of a genetic nucleus. The number of pigs per pen is approximately four, seven, and thirteen for phases one to three, respectively. The quarantine nursery was not available for cycle 1 (first seven batches), for which phases 1 and 2 were combined. During this period, strict biosecurity was practiced between the nursery and finishing unit (same building connected by a hallway) but this was not sufficient to keep diseases from getting into the nursery, after which the quarantine nursery was established.

The natural disease challenge was established by bringing in naturally infected animals (seeder pigs) from strategically selected farms into the challenge barn (late nursery and finishing). Four groups of 12–28 pigs were introduced from three different commercial farms in the first four months of the study as seeder pigs. Thereafter, monitoring for diseases was focused on the test population and less on the seeder pigs. Initially, the targeted diseases included porcine reproductive and respiratory syndrome virus (PRRSV), porcine circovirus type 2 (PCV2), *Mycoplasma hyopneumoniae* (*M. hyo.*), *Actinobacillus pleuropneumoniae* (APP), and swine influenza, and various opportunistic bacterial pathogens, including *Streptococcus suis* and *Haemophilus parasuis*. APP strain 12 was present. Three different strains of PRRSV present had ~85–90% sequence identity to the PRRS-MLV (Boehringer Ingelheim, St. Joseph, MO). Every batch was confirmed to have been exposed to PRRSV based on sampling a subset of individuals using PCR and serology four- and six-weeks post challenge, respectively. Multiple influenza subtypes were present in the barn including the H1N1 and H3N2 based on serological testing of a subset of the population at 18 weeks post entry. No typing for PCV2 or *M. hyo* was completed. The disease challenge was a function of these pathogens collectively in combination with the environment, management, and veterinary strategies designed to obtain a target infection pressure for each batch. The natural challenge was set up as a continuous flow system in order to maintain a steady health challenge without having to keep introducing pathogens, as well as for labor and flow considerations. A new batch of naïve pigs enters every three weeks and is generally provided fence-line contact with the preceding batch for ~1-week period, except during periods of excessively high infection pressure when it is discontinued to help reduce mortality rate to sustainable levels

established by the Animal Protection Committee. For the data used in this study, the following viruses were identified in the challenge facility: PRRSV (3 strains), Influenza A virus of swine (AIV; 2 strains), porcine circovirus type-2 (PCV2), and porcine rotavirus A (RVA). Bacterial pathogens diagnosed included: *Actinobacillus pleuropneumoniae* (APP), *M. hyo.*, *Streptococcus suis*, *Haemophilus parasuis*, *Brachyspira hampsonii*, *Salmonella sp.*, *Cystoisospora suis* (Coccidiosis), *Ascaris suum*, *Erysipelothrix rhusiopathiae*, and *Staphylococcus hyicus* (causative agent for Exudative Epidermitis). Not all pathogens were identified in all batches, as would be the case on a commercial farm and other unidentified minor pathogens may also have been present. Although fairly endemic in the US, porcine epidemic diarrhea (PED) was not present in Québec and was therefore not present in the challenge facility.

To maintain acceptable levels of animal welfare and morbidity, individual treatments were given on a case-by-case basis, along with periodic batch-level (or mass) treatments. The treatment protocol was established by the consulting veterinarian, who is licensed in the province of Québec, Canada. Veterinarians had close oversight on the treatment protocol over time, which was adapted as needed to maintain acceptable levels of disease and minimize animal suffering. In addition, some treatment decisions were made by multiple veterinarians and trained barn staff, introducing some level of subjectivity, as would be the case in a commercial facility. Pigs exhibited clinical signs indicative of pneumonia, diarrhea, lameness, arthritis, meningitis, dermatitis, pallor, lethargy, weight loss, unthriftiness, cyanosis, or conjunctivitis. Pigs were treated with one of ten different antibiotics as per a regimented treatment protocol outlining primary and secondary (if needed) treatment choices for each ailment. For some clinical signs, one of two anti-inflammatory drugs were also administered. Batch-level water medication was used in the nursery when deemed necessary during periods of severe illness. One of two antibiotics were used in these batches. Furthermore, a water-soluble anti-inflammatory drug was also periodically administered in the nursery to treat batches that suffered from severe respiratory disease (primarily related to PRRSV infection). After the first seven batches, vaccination for PCV2 was added to the quarantine protocol in response to necropsy data linking characteristic lymphoid lesions with the presence of the virus. Reports from feed intake recording were generated daily for farm staff to identify sick pigs that did not eat as much as expected. Euthanasia decisions for animal welfare reasons were made by farm staff, with appropriate veterinary oversight. Barn air and temperatures were controlled with a ventilation system and a heater was used to regulate the lower bound temperatures within the barn.

A new batch of pigs entered the natural challenge protocol every three weeks. Each batch consisted of ~60 or ~75 weaned Large White by Landrace (or reciprocal mating) barrows (castrated male pigs) that were provided by one of the seven members of PigGen Canada (<https://piggenCanada.org/>) from healthy multiplier farms. Each batch was sourced from one multiplier, but over time different multipliers could supply pigs for a given PigGen member. Variables collected on piglets at the multiplier farms were date of birth, wean age, and biological

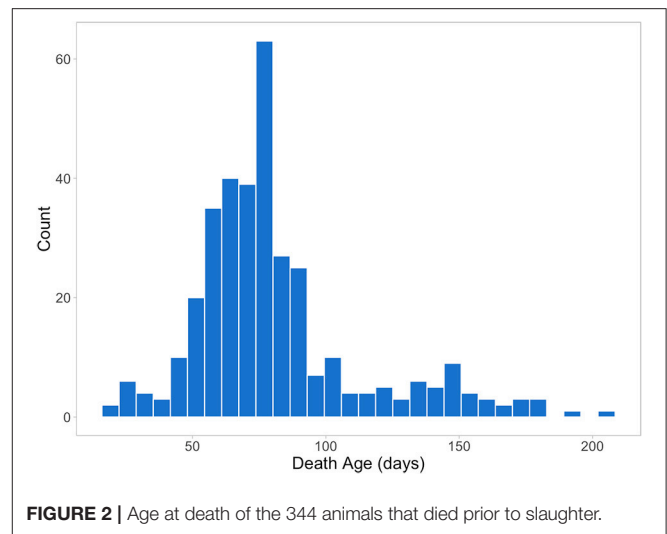


sow ID. The protocol specified that two to four weaned barrows should be sampled per litter. Eighty-seven percent of all piglets met that criterion. Piglets were retagged with a sequential ID tag when they arrived at the first nursery. Every seven batches were considered a cycle, numbered one to three in the current study. Each company was represented once each cycle (i.e., one batch per company per cycle). This continued for a total of three cycles, therefore each company was represented three times in the data analyzed here. This came to a total of 1,341 pigs that entered the facilities within the time period studied.

A fixed weight system was used to identify pigs for slaughter, starting at ~180 days of age. Pigs that were not heavy enough were delayed for three weeks and then evaluated again. Most batches took between two to four slaughter groups to slaughter all pigs from a batch. **Figure 1** shows a timeline of all batches analyzed in this study, with timing of date of birth, entry into the first nursery, entry into the finisher, and slaughter dates.

Data

All data and samples were collected by trained research staff from CDPQ following established protocols. Body weights were taken approximately every three weeks. However, if a pig was unhealthy, it may have been weighed a few more times, closer together in time. To obtain daily weights, a LOESS (Locally Weighted Scatterplot Smoothing) regression was fit to all weights available for an animal, using the loess function in R using defaults (R core team, 2017). LOESS regression is a form of nonparametric regression, also known as local regression, that can fit non-linear trends in a flexible enough manner to “connect the dots” between weight measurements. The correlation of LOESS predicted with observed weights was 0.9995 for days with an observed weight. The LOESS predicted daily weights were utilized for calculations of production measures such as feed efficiency and growth (see below).

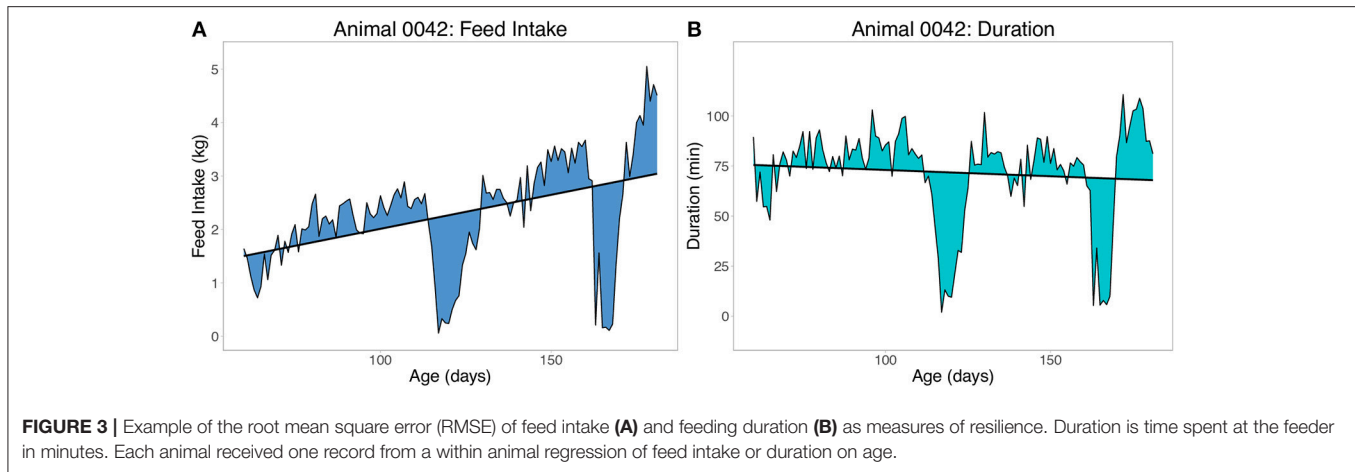


Feed intake data was recorded in the finishing phase using IVOG[®] feeding stations (Insentec, Marknesse, Netherlands). Feed was available *ad libitum* throughout the study. The nursery feeding protocol consisted of four diet phases, while the finishing period included two diet phases. Individual feed intake visits were processed and cleaned by CDPQ staff using the methods of Casey et al. (2005) and were aggregated into daily totals for each pig, including total amount of feed consumed (kg) and duration (time) at the feeder (minutes). Daily totals of more than 5 kg of feed were set to missing. Missing daily values were subsequently imputed using a 5-day rolling average within animal (also used if there were two adjacent days missing).

Figure 2 shows the distribution of death age for pigs that died prior to slaughter (344 or 26% of the 1,341 total animals). All treatment and mortality events and reasons were recorded (assigned by CDPQ research staff). Main treatment reasons included respiratory distress (thumping), gray/brown scours, coughing, lameness, yellow scours, arthritis, and failure to thrive/poor/skinny/hairy. Main mortality reasons included failure to thrive/poor/skinny/hairy, thumping/heavy breathing, sudden death, meningitis, and lameness/arthritis. Only individual treatments were included in the analyses and batch treatments were removed. Virtually all treatment reasons and ~89% of the mortality reasons were disease-related.

Traits

Traits used for validation of the resilience traits developed herein included mortality (binary 0/1, 1 = died), number of treatments (TRT), and number of treatments per 180 days (TRT180). Number of treatments was a count of the number of individual treatments received by a pig. An individual treatment included any drug injection into an individual animal. Group treatments applied to batches were not included, as these would be accounted for in the model by the fixed effect of batch anyway. Only pigs that survived to slaughter received a phenotype for TRT. TRT180 was the number of treatments standardized to 180 days and was computed for animals that reached 65 days of age



(approximate age of entry into the finishing unit). For instance, if an animal received three individual treatments and died on day 80, the animal's adjusted TRT180 was $(3/80) \times 180 = 6.75$. This was to standardize treatment rate to approximately the same scale as TRT and to be interpretable from a practical standpoint (number of treatments to slaughter).

Two sets of resilience traits were derived from the daily FI data available for each pig. The first set of traits were derived as the root mean square error (RMSE) within animal from the regression of feed intake (FI) or duration (DUR) on age (RMSE_{FI} and RMSE_{DUR}, respectively), using ordinary least squares (OLS) linear regression. Duration is the daily time spent at the feeder in minutes. An example of the RMSE for one pig with two large deviations from illness is shown in **Figure 3** for FI (**Figure 3A**) and duration (**Figure 3B**). To obtain a phenotype for RMSE, animals had to have a minimum of 60 days of FI recorded. A less resilient animal is expected to have a larger value for RMSE. Preliminary analyses showed that without setting this minimal number of days, animals that died early in finishing were grouped on the left side of the distribution of RMSE (i.e., they would be considered more resilient). Duration (time) at the feeder was chosen over traits such as number of meals due to its strong association with off-feed events (e.g., **Figure 3**).

The second set of novel resilience phenotypes was based on quantile regression (QR), which can be useful for regression problems that include heterogeneous variances (Cade and Noon, 2003). A 5% quantile regression was performed using all data across batches, separately for FI and duration (**Figures 4A,B**). Negative residuals (below the regression line) from these regression equations were used to classify a day of FI or duration for an individual pig as an off-feed day (**Figures 4C,D**). These were aggregated within animal to a proportion of "off-feed" days (one record per animal). As with RMSE, each animal received only one phenotype for FI and for duration (QR_{FI} and QR_{DUR}). The 5% threshold was set based on **Figures 4A,B**, as it separated the "cloud" of relatively healthy days from off-feed days, as well as appraisal of FI plots within animal. In total, 258 animals (25%) did not have any day below the 5% quantile regression for QR_{FI}, while a 1% threshold resulted in 677 animals (65%) not having

any days below the threshold. To obtain a phenotype for QR, animals had to have at least 60 days of FI recorded (same for RMSE). As with RMSE, susceptible animals are expected to have larger values for QR than resilient animals.

Production traits analyzed included nursery ADG (NurADG), finishing ADG (FinADG), average daily feed intake (ADFI), feed conversion ratio (FCR), residual feed intake (RFI), carcass weight (CWT), dressing proportion (DRS), lean yield (LYLD), carcass backfat (CBF), and carcass loin depth (CLD). To obtain a phenotype for a production trait, pigs had to complete the corresponding phase (nursery or finishing). Nursery and finishing ADG were calculated from regression slopes of daily LOESS weights (see above) on age for the entire nursery period (quarantine and challenge nursery) and the finishing period, respectively. NurADG started when the pig entered the quarantine nursery. NurADG ended and FinADG started the first day FI was recorded. LOESS predicted daily weights were used to compute ADG because a weight was not always available for the days when pigs were moved to the finishing unit. Also, some animals received more weights prior to being euthanized or death, which would influence the regression of weight on age (not evenly spaced). The impact of using LOESS predicted instead of observed weights was very small for FinADG (more weights) but was more significant for NurADG, as the correlation for FinADG with or without use of LOESS prediction was very high when using the closest endpoints, but much lower for NurADG. This was because the nursery period was much shorter, and many pigs only had two weights prior to being moved to the finishing unit and therefore a larger change in ADG was observed. Average daily feed intake (ADFI) was the average feed intake of daily records during the finishing period. Feed conversion ratio (FCR) was defined as the sum of daily records for FI over the total body weight gain for that same finishing period. Residual feed intake (RFI) was computed in a one-step analysis following Cai et al. (2008), using ADFI as the response variable and average body weight (average weight in the finisher), finishing ADG, and ultrasound backfat as covariates, along with other fixed effects, as described below. Ultrasound backfat was taken just prior to slaughter at the 10th rib. Dressing proportion was calculated

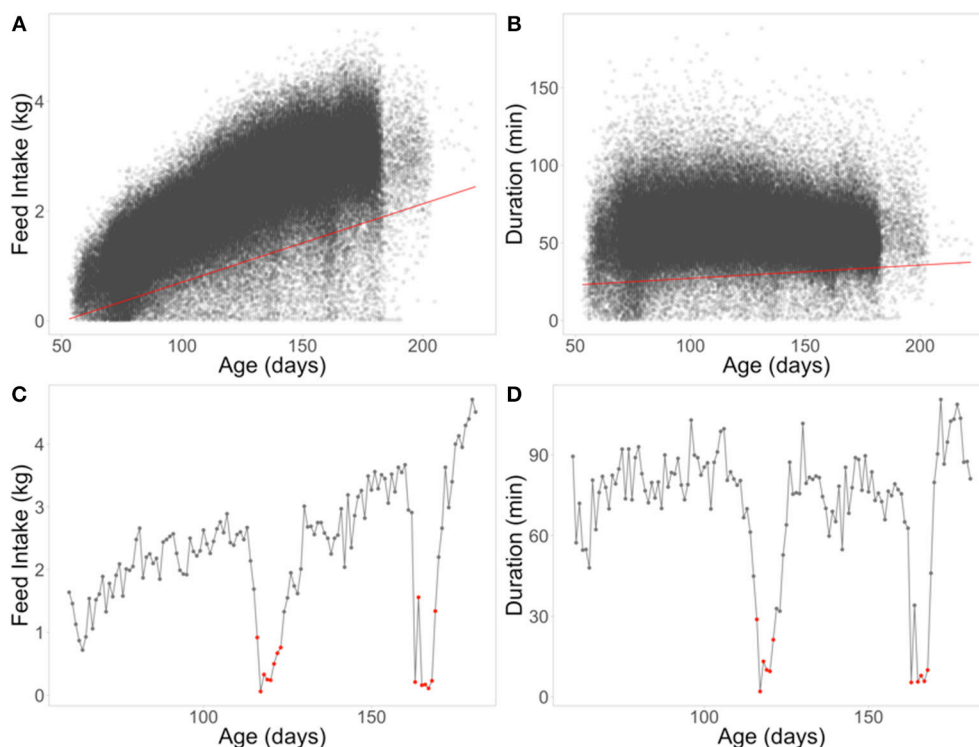


FIGURE 4 | Quantile regression (QR) at the 5% level for feed intake (A) and duration (B). Example of classifying off-feed days for animal 0042 using QR_{FI} (C) and QR_{DUR} (D) in the lower panels. Red lines are the 5% QR for each FI (A) and duration (B). Red dots represent off-feed days used to calculate the QR phenotypes (proportion of off-feed days).

by dividing the carcass weight (head on, leaf lard in, warm carcass) by the live weight prior to slaughter. Carcass backfat (CBF) and loin depth (CLD) were recorded using a Destron FearingTM machine (Texas, USA) at the abattoir. Lean yield was calculated using the following regression equation for lean yield in Québec: $LYLD = 68.1863 - (0.7833 \cdot CBF) + (0.0689 \cdot CLD) + (0.008 \cdot CBF^2) - (0.0002 \cdot CLD^2) + (0.0006 \cdot CBF \cdot CLD)$ (Pomar and Marcoux, 2003). This equation was mostly driven by backfat ($r = -0.98$). Not all batches had carcass data, leading to some variation in the number of observations for these traits. Carcass phenotypes were also captured at different time points within batch due to the protocol to only send the pigs that met market weight at each slaughter date, as mentioned above. The average live weight at slaughter was 118.9 kg.

Genotyping

Animals were genotyped with the 650k Affymetrix Axiom Porcine Genotyping Array by Delta Genomics (Edmonton AB, Canada). In total, 658,692 single nucleotide polymorphisms (SNPs) were included on the chip. Raw Affymetrix SNP data output was processed separately for each cycle by Delta Genomics with the Axiom[®] Analysis Suite using all defaults. The SNPs that passed quality control for all three cycles were utilized for analysis, for a total of 516,066 SNPs. Imputation of missing genotypes was completed with FImpute (Sargolzaei et al., 2014). The pedigree was utilized for imputation but only included the

dam at the multiplier, since sire was typically unknown due to the use of pooled semen. Genotypes were then processed using the preGSf90 software from the BLUPF90 family of programs, using defaults (Misztal et al., 2002). Genotypes on seventeen samples were found to be duplicates and were removed. After all quality control, genotypes on 1,215 animals and 487,762 SNPs remained.

Variance Component Analyses

Variance components were estimated by single-step GBLUP with the **H** matrix (Legarra et al., 2009; Christensen and Lund, 2010), using the BLUPF90 family of programs (Misztal et al., 2002). Data included phenotypes on 1,341 animals, of which 1,215 had genotypes. Basic animal models were fit for all traits, with random animal genetic effects (using the **H** matrix) and random residuals. The genomic relationship matrix (**G**) was calculated using $ZZ'/\text{sum}(2pq)$ (VanRaden, 2008), where **Z** = **M-P**. Only the dam was available to construct the **A** matrix. Single trait models were used to obtain heritability estimates and bivariate models for genetic correlations. Models for mortality and number of treatments included batch and age of entry into the quarantine nursery as fixed effects and were modeled as linear traits. Mortality was initially analyzed as a threshold trait but resulted in unrealistically large estimates of heritability. A simulation was used to confirm that threshold models tended to significantly overestimate heritability with small sample sizes such as this study. One alternative could be to use a more

TABLE 1 | Counts and means for measures of resilience in three cycles of the natural challenge experiment ($n = 1,341$ total animals entered).

| Cycle ^a | Count | Mortality, % | TRT ^a | TRT180 ^b | RMSE _{FI} ^c , kg | RMSE _{DUR} ^c , min | QR _{FI} ^d | QR _{DUR} ^d |
|--------------------|-------|--------------|------------------|---------------------|--------------------------------------|--|-------------------------------|--------------------------------|
| 1 | 441 | 35 | 1.45 | 2.63 | 0.48 | 13.90 | 0.06 | 0.05 |
| 2 | 452 | 13 | 1.96 | 2.07 | 0.46 | 11.90 | 0.04 | 0.05 |
| 3 | 448 | 29 | 1.89 | 2.61 | 0.46 | 13.40 | 0.04 | 0.04 |

^aNumber of treatments, animals must have made it to slaughter.

^bTreatment rate adjusted to 180 days, animals must have made it through 65 days of age to obtain a phenotype.

^cRoot mean square error (RMSE) from the within animal regression of Feed Intake (FI) or Duration (DUR) on age with at least 60 days of FI.

^dQuantile regression (QR) from using the 5% QR over all the feed intake (FI) or duration (DUR) data and then aggregating off-feed days within animal as a proportion.

recent approach from Ødegård et al. (2010) but mortality was not the main focus of this research. More data may be needed to analyze mortality as a threshold trait. Analyses for finishing traits included fixed effects of batch, finishing start age, and finishing pen. Litter effects (random) were minimal (below 0.05 for the proportion explained and within one SE of zero) for the traits analyzed and, therefore, were subsequently dropped from all analyses. Litter effects were also difficult to estimate, with an average of 2.02 litter mates per pig. Not all animals survived to record a phenotype for traits recorded later such as FinADG or carcass traits, therefore for these traits the average was <2 litter mates per pig.

RESULTS

Table 1 shows summary statistics for the three cycles of data used in the analyses (seven batches per cycle). Batches included from 59 to 77 pigs, except for one batch of 28 (not shown), and each cycle ranged from 441 to 452 pigs (1,341 total). Mortality was highest in cycle one (35%), decreased in cycle 2 (13%), and then returned to a higher rate in cycle 3 (29%). Mortality per batch ranged from 4 to 57%, with the median being 18%. The continuous flow system maintained pathogen burden throughout the study, however, seasonality clearly led to higher mortality during the winter months. In contrast to TRT180, TRT did not follow the mortality trend due to the requirement of survival to slaughter. In general, the FI resilience phenotypes followed the same time trend as mortality, except for QR_{DUR}.

Table 2 shows the number of observations and summary statistics for each trait. RMSE and QR measures of resilience were required to have 60 days of FI to receive a phenotype, which removed 188 animals from those that made it into the finishing unit. The average RMSE_{FI} was 0.47 kg, ranging from 0.19 to 0.97 kg. RMSE_{DUR} averaged 13.10 min, with a range of 5.71 to 37.54 min. One major difference between TRT and TRT180 was that TRT180 allowed animals that died after 65 days of age to record a phenotype, which added 219 phenotypes. Of those that survived, the number of treatments was 1.79 on average, but 2.43 for TRT180 (median of 1.97). Due to the health challenges, many of the production phenotypes had a wide range. Nursery ADG ranged from 0 to 0.67 kg/d and finishing ADG from 0.36 to 1.20 kg/d. This caused carcass weights to have a wide range as well, despite the aim to slaughter at a “fixed weight.”

Table 3 shows estimates of heritabilities and genetic correlations among the resilience traits and between resilience

traits and production traits. Many estimates had large SE due to relatively small sample sizes. Heritability estimates for the novel resilience traits ranged from 0.15 to 0.26. Mortality had a heritability estimate of 0.13 ± 0.05 , while TRT and TRT180 had estimated heritabilities of 0.13 ± 0.07 and 0.29 ± 0.07 , respectively. The estimate of the genetic correlation between mortality and TRT180 was 0.93 ± 0.29 (results not shown). Estimates of genetic correlations among the novel resilience measures ranged from 0.01 to 0.67, indicating they are different genetic traits. Estimates of genetic correlations of mortality and TRT180 with novel resilience traits were positive, as expected, and ranged from 0.37 to 0.85. Due to data processing and removal of phenotypes from TRT because of the requirement of survival to slaughter, TRT180 was deemed to be a better phenotype for validation of the novel traits (**Table 2**). The estimate of the genetic correlation of RMSE_{DUR} was 0.12 ± 0.76 with TRT and 0.62 ± 0.13 with TRT180. Of the two RMSE measures of resilience, RMSE_{DUR} was more highly correlated genetically with mortality and treatments than RMSE_{FI}. For the QR traits, QR_{FI} had a slightly higher genetic correlation with mortality and number of treatments than QR_{DUR}, which could be because farm staff received daily reports of which pigs were not eating enough feed and were flagged for further evaluation (see discussion).

Estimates of genetic correlations of RMSE traits with production traits were low, but many were favorable (**Table 3**). Nursery ADG was unfavorably correlated with RMSE_{FI} (0.77 ± 0.24) but most of the other production traits had favorable or close to zero genetic correlations with the two RMSE measures of resilience. Finishing ADG had a genetic correlation estimate of -0.31 ± 0.26 with RMSE_{FI} and of -0.19 ± 0.26 with RMSE_{DUR}. Feed efficiency based on FCR and RFI were genetically correlated with RMSE_{FI} (0.39 ± 0.21 and -0.22 ± 0.27 , respectively). Resilience based on QR measures was more strongly associated with production traits than resilience based on RMSE. Both QR_{FI} and QR_{DUR} had strong genetic correlations with FinADG, at -0.75 ± 0.26 and -0.70 ± 0.17 , respectively. Notice, however, that QR was not strongly correlated with NurADG, likely because feed intake was only collected in the finisher. ADFI was negatively correlated with QR_{FI} and QR_{DUR}, at -0.79 ± 0.19 and -0.58 ± 0.16 , respectively. Estimates of genetic correlations of QR with FCR were low, at -0.14 ± 0.35 and 0.02 ± 0.24 , vs. -0.78 ± 0.21 and -0.63 ± 0.16 with RFI, which were similar to those for ADFI. Carcass BF and LD had negative genetic correlations with QR measures of resilience (-0.36 to -0.21).

TABLE 2 | Summary statistics of the analyzed traits ($n = 1,341$ total animals).

| Trait ^a | Number of phenotypes | Mean | SD | Median | Min | Max |
|---------------------------|----------------------|-------------------|-------|-------------------|-------------------|-------------------|
| Mortality | 1341 | 0.26 ^b | N/A | 0.18 ^c | 0.04 ^c | 0.57 ^c |
| TRT ^d | 997 | 1.79 | 1.56 | 1 | 0 | 10 |
| TRT180 ^d | 1216 | 2.43 | 2.33 | 1.97 | 0 | 15.32 |
| RMSE _{FI} , kg | 1036 | 0.47 | 0.11 | 0.45 | 0.19 | 0.97 |
| RMSE _{DUR} , min | 1036 | 13.10 | 4.36 | 12.30 | 5.71 | 37.54 |
| QR _{FI} | 1036 | 0.04 | 0.07 | 0.01 | 0 | 0.67 |
| QR _{DUR} | 1036 | 0.04 | 0.06 | 0.02 | 0 | 0.52 |
| NurADG, kg/day | 1218 | 0.32 | 0.12 | 0.31 | 0.00 | 0.67 |
| FinADG, kg/day | 992 | 0.89 | 0.14 | 0.91 | 0.36 | 1.20 |
| ADFI, kg/day | 997 | 2.20 | 0.33 | 2.23 | 0.97 | 3.01 |
| FCR, kg/kg | 997 | 2.58 | 0.21 | 2.57 | 1.94 | 3.34 |
| RFI, kg | 991 | N/A ^e | | | | |
| CWT, kg | 837 | 93.77 | 10.56 | 95.00 | 49.90 | 118.60 |
| DRS | 837 | 0.78 | 0.02 | 0.78 | 0.68 | 0.84 |
| LYLD | 799 | 60.92 | 1.71 | 60.90 | 55.20 | 65.60 |
| CBF, mm | 800 | 17.96 | 3.87 | 17.50 | 7.50 | 33.50 |
| CLD, mm | 800 | 60.69 | 6.14 | 60.50 | 41.50 | 81.00 |

^aTRT, number of treatments for animals that made it to slaughter; TRT180, treatment rate adjusted to 180 days for pigs that made it to 65 days of age; RMSE, root mean square error (novel phenotype with FI or duration); QR, quantile regression as a proportion (novel phenotype with FI or duration); NurADG, nursery ADG; FinADG, finishing ADG; ADFI, average daily feed intake; FCR, feed conversion ratio (kg feed / kg weight gain); RFI, residual feed intake (adjusted for FinADG, metabolic weight, and ultrasound backfat); CWT, carcass weight; DRS, dressing proportion; LYLD, lean yield (equation using backfat and loin depth); CBF, carcass backfat; CLD, carcass loin depth. RMSE and QR phenotypes required 60 days of FI.

^bOverall mortality proportion.

^cMedian, min, and max by batch, not individual.

^dTRT required the animal to survive to slaughter. TRT180 required the animal survive to 65 days of age.

^eResidual feed intake (RFI) was calculated using ADFI as the response in a one-step method.

Table 4 shows estimates of genetic correlations of production traits with mortality and number of treatments. Production traits tended to have low genetic correlations with mortality (<0.30 in absolute value) but higher with number of treatments for some traits. Estimates of the genetic correlation of finishing ADG and ADFI with TRT and TRT180 ranged from -0.60 to -0.70 . Carcass weight also showed a strong negative genetic correlation of -0.67 ± 0.14 with TRT180, similar to FinADG. Carcass BF, LD, and LYLD were weakly genetically correlated with both number of treatments and mortality.

DISCUSSION

Novel disease resilience measures were extracted from daily feed intake data of grow-finish pigs that were exposed to a multifactorial natural disease challenge that was designed to mimic a commercial environment with high disease pressure to maximize the expression of genetic differences of resilience between animals. Although the specific disease and environmental conditions that were established in this study cannot be exactly replicated, the general protocols established can

be replicated in both research and commercial settings, similar to the replication of field studies on health-challenged farms. Moreover, although infection pressure waxes and wanes over time, it is assumed to be relatively consistent within batch because of the close proximity in which new batches are housed.

The resilience traits that were derived from individual daily feed intake data showed moderate heritabilities and moderate to strong genetic correlations with mortality and treatment rate. Genetic correlations production traits tended to be low for the RMSE measures of resilience but higher for the QR measures. Data from the most important disease exposure period, i.e., the challenge (2nd) nursery, were not included in either RMSE or QR measures of resilience because individual feed intake could only be collected in the finishing unit. The challenge nursery period was, however, critical, as this represented the first exposure to many pathogens in the barn for most batches (nose-to-nose contact for new batches with older already infected batches). Thus, pigs could have been infected with pathogens and recovered in the nursery before feed intake recording started in the finishing unit. This may have reduced genetic correlations of the evaluated novel resilience traits with mortality or number of treatments. Future research could address this by collecting important phenotypic data during the entire challenge period or by setting up the nursery away from the finishing challenge facility.

The RMSE measures of resilience proposed here were designed to quantify severity of disease and other stressors on individual animals over time (see below), whereas QR measures of resilience classified days as off-feed events, reflecting more extreme events, making the QR measures less sensitive and showing less variation than RMSE measures. This may partially explain the slightly lower estimates of heritability for QR compared to RMSE measures and the higher genetic correlations of QR with TRT180 and mortality than RMSE, as both mortality and treatments are the result of severe clinical disease. Pigs were typically not euthanized until the disease had progressed and the animal was clearly suffering. Treatments were generally given only when clinical signs of illness were present (e.g., diarrhea, coughing, lethargy, etc.). RMSE measures of resilience may have the ability to capture subclinical disease and other stresses in addition to clinical disease, enabling it to be more sensitive than number of treatments, mortality, and QR measures of resilience, which typically capture only severe events. This would make RMSE measures of resilience different traits than treatments, mortality, and QR, which was supported by the estimates of genetic correlations. Although QR measures of resilience can also capture the effects of stressors other than disease, it is less likely to do so compared to RMSE due to the larger impact of disease on feed intake compared to other stressors (results not shown).

Quantile regression measures of resilience tended to have higher genetic correlations with production traits than RMSE, likely because pigs that grow slower typically have lower ADFI and, thus, when they get sick, they need a smaller drop in FI to drop below the QR line. In contrast, animals with high ADFI must drop further to have a drop below the QR threshold. Thus, pigs with low average FI across the finishing period are expected to have more days classified as being off-feed days, resulting in

TABLE 3 | Estimates of heritability (SE) for traits analyzed and of genetic correlations (SE) with resilience measures ($n = 1341$ total animals, see **Table 2** for actual counts per phenotype).

| Trait ^a | h^2 | Genetic correlation (SE) with | | | |
|---------------------------|-------------|-------------------------------|---------------------|------------------|-------------------|
| | | RMSE _{FI} | RMSE _{DUR} | QR _{FI} | QR _{DUR} |
| RMSE _{FI} , kg | 0.21 (0.07) | – | 0.47 (0.26) | 0.50 (0.31) | 0.52 (0.24) |
| RMSE _{DUR} , min | 0.26 (0.07) | | – | 0.67 (0.28) | 0.01 (0.29) |
| QR _{FI} | 0.15 (0.06) | | | – | 0.64 (0.30) |
| QR _{DUR} | 0.23 (0.07) | Symmetric | | | – |
| Mortality | 0.13 (0.05) | 0.37 (0.34) | 0.60 (0.26) | 0.75 (0.27) | 0.70 (0.21) |
| TRT | 0.13 (0.07) | 0.52 (0.48) | 0.12 (0.76) | 0.76 (0.58) | 0.62 (0.56) |
| TRT180 | 0.29 (0.07) | 0.56 (0.18) | 0.62 (0.13) | 0.85 (0.16) | 0.65 (0.15) |
| NurADG, kg/day | 0.45 (0.07) | 0.77 (0.24) | –0.10 (0.19) | –0.11 (0.25) | 0.21 (0.20) |
| FinADG, kg/day | 0.25 (0.07) | –0.31 (0.26) | –0.19 (0.26) | –0.75 (0.26) | –0.70 (0.17) |
| ADFI, kg/day | 0.32 (0.07) | 0.03 (0.26) | –0.24 (0.21) | –0.79 (0.19) | –0.58 (0.16) |
| FCR, kg/kg | 0.35 (0.07) | 0.39 (0.21) | –0.17 (0.25) | –0.14 (0.35) | 0.02 (0.24) |
| RFI, kg | 0.24 (0.07) | –0.22 (0.27) | –0.35 (0.25) | –0.78 (0.21) | –0.63 (0.16) |
| CWT, kg | 0.31 (0.08) | –0.04 (0.28) | –0.13 (0.24) | –0.78 (0.25) | –0.63 (0.17) |
| DRS | 0.10 (0.06) | –0.23 (0.07) | –0.49 (0.49) | –0.73 (0.60) | –0.52 (0.53) |
| LYLD | 0.50 (0.08) | 0.13 (0.24) | 0.00 (0.23) | 0.50 (0.24) | 0.37 (0.19) |
| CBF, mm | 0.46 (0.09) | –0.14 (0.26) | 0.03 (0.23) | –0.36 (NA) | –0.35 (0.18) |
| CLD, mm | 0.39 (0.08) | –0.20 (0.27) | –0.05 (0.24) | –0.29 (0.30) | –0.21 (0.25) |

^a RMSE, root mean square error (for FI or duration); QR, quantile regression as a proportion (for FI or duration); TRT, number of treatments for animals that made it to slaughter; TRT180, treatment rate adjusted to 180 days for pigs that made it to 65 days of age; NurADG, nursery ADG; FinADG, finishing ADG; ADFI, average daily feed intake; FCR, feed conversion ratio (kg feed / kg weight gain); RFI, residual feed intake (adjusted for FinADG; metabolic weight; and ultrasound backfat); CWT, carcass weight; DRS, dressing proportion; LYLD, lean yield (equation using backfat and loin depth); CBF, carcass backfat; CLD, carcass loin depth. RMSE and QR phenotypes required 60 days of FI.

higher genetic correlations of QR measures of resilience with traits that are closely related to FI such as ADG, than RMSE. Refining these resilience phenotypes will be a focus of future research.

Feed Intake Duration

Feeding duration was used in this study as a proxy for drops in FI. In the past, there have been many attempts to link feeding traits with FI (de Haer et al., 1993; Von Felde et al., 1996; Young et al., 2011; Lu et al., 2017). In animal breeding, these traits include duration (time at the feeder), number of visits, and feed intake rate. Previous studies were typically conducted in healthy environments and feeding traits such as duration at the feeder may become more valuable under disease challenge. **Figures 3, 4** show how the pattern of FI and duration were very similar across time for this selected animal. Measures of resilience based on duration had comparable genetic correlations with mortality and number of treatments as measures of resilience based on FI in the present study. Day-to-day variation in duration at the feeder could be more applicable on commercial farms if commercial feeders could be retrofitted to record individual time at the feeder using antennae and RFID tags. This could also be extended into the nursery, allowing feeding traits to be collected over the entire wean-to-finish period. Additional research is needed to evaluate other feeding traits that can be extracted from electronic feeders (Kyriazakis and Tolkamp, 2018). Feeding patterns within a day may be useful and could be utilized to better quantify resilience.

The current study took the simple approach and used daily totals, but this is only a starting point for more research on this topic.

Causes of Variation in FI and Their Relationship With Resilience

Colditz and Hine (2016) presented a holistic view of resilience by including other stressors to define general environmental resilience. In the current study, it is not possible to verify that all drops in FI and duration at the feeder observed in our data are due to disease alone. Martínez-Miró et al. (2016) categorized animal stressors into social, environmental, metabolic, immunological, and human interactions. Each of these could be decomposed into more detailed stressors. For instance, immunological stressors can be broken down further into individual resistance, tolerance, or resilience toward PRRSV or PCV2 (among others). There can also be interactions between these stressors (Salak-Johnson and McGlone, 2007), although other studies have suggested some stressors may be additive (Hyun et al., 1998).

There is a long list of stressors that can impact feed intake and performance on swine. The impact of pathogens on feed intake has been well established in the literature (Sandberg et al., 2006; Kyriazakis and Doeschl-Wilson, 2009) and is dependent upon, but is not limited to, the type of pathogen, the strain of the pathogen, previous exposure, and vaccinations. Porcine reproductive and respiratory syndrome virus alone costs the swine industry an estimated \$664 million annually in the US

TABLE 4 | Estimates of genetic correlations (SE) of mortality and number of treatments with production traits.

| Trait ^a | Mortality | TRT | TRT180 |
|--------------------|------------------|------------------|--------------|
| NurADG, kg/day | 0.27 (0.44) | −0.33 (1.28) | −0.06 (0.16) |
| FinADG, kg/day | −0.06 (0.36) | −0.68 (0.42) | −0.70 (0.13) |
| ADFI, kg/day | −0.04 (0.28) | −0.60 (0.32) | −0.62 (0.13) |
| FCR, kg/kg | 0.24 (0.28) | −0.15 (0.43) | 0.13 (0.18) |
| RFI, kg | −0.29 (0.31) | −0.41 (0.45) | −0.53 (0.19) |
| CWT, kg | 0.02 (0.33) | −0.57 (0.36) | −0.67 (0.14) |
| DRS, | dnc ^b | dnc ^b | −0.63 (0.35) |
| LYLD, | 0.01 (0.34) | −0.14 (0.40) | −0.01 (0.20) |
| CBF, mm | dnc ^b | 0.17 (0.48) | 0.01 (0.21) |
| CLD, mm | 0.27 (0.33) | −0.04 (0.38) | −0.12 (0.20) |

^aNurADG, nursery ADG; FinADG, finishing ADG; ADFI, average daily feed intake; FCR, feed conversion ratio (kg feed/kg weight gain); RFI, residual feed intake (adjusted for FinADG, metabolic weight, and ultrasound backfat); CWT, carcass weight; DRS, dressing proportion; LYLD, lean yield (equation using backfat and loin depth); CBF, carcass backfat; CLD, carcass loin depth.

^bDid not converge.

(Holtkamp et al., 2013). Heat stress is another common reason why animals deviate from their expected FI (Guy et al., 2017), which has been characterized in growing pigs (Rauw et al., 2017b) and in sows (Vilas Boas Ribeiro et al., 2018). Mycotoxins have been known for a long time to influence feed intake (Smith et al., 1997). Social interactions (i.e., space requirements) are another common source of stress in swine (Hyun et al., 1998). These social effects have been investigated in piglets (Bouwman et al., 2010), in growing pigs (Street and Gonyou, 2008), as well as in sows (Hemsworth et al., 2013). Martínez-Miró et al. (2016) discussed many other stressors including human handling, vaccination, dust/gas/ammonia, and out of feed and water events.

Knap (2009) originally used an example of heat stress in pigs to show the potential relevance of day-to-day variability in feed intake. The measures of resilience developed here could also be used to quantify resilience to heat tolerance (Fragomeni et al., 2016; Guy et al., 2017), activity level (Sadler et al., 2011; Gilbert et al., 2017; King et al., 2018), and possibly even reduce stressful interactions for pigs (Rauw et al., 2017a). Heat stress was estimated to cost the US swine industry \$299 million per year (St-Pierre et al., 2003). These measures could also be based on other sources of data such as water intake data (Madsen and Kristensen, 2005; Rusakovica et al., 2017) or body temperature recordings on individual pigs (Petry et al., 2005, 2017). Elgersma et al. (2018) developed variation and “drop phenotypes” from milk yield data in dairy cows. The phenotypes developed in the current study could also be used to develop similar phenotypes for other species.

A problem with the interpretation of the types of resilience measures developed here and by Elgersma et al. (2018) is that factors influencing resilience phenotypes in general are still a “black-box” (Mulder and Rashidi, 2017), not only in terms of different diseases but for all the other stressors described above. This is one reason why we cannot expect the genetic correlation between RMSE and mortality or treatments to be one, as factors that influence feed intake, could be non-health

related. Another reason may be that RMSE captures sub-clinical disease better than QR (Elgersma et al., 2018 mentions this also for their resilience traits). Although from a practical or commercial breeding standpoint, it probably matters little why animals deviate from expected feed intake. Traits presented in the current study should be thought of as having economic value (Elgersma et al., 2018). The usefulness of these novel traits in a breeding program will depend on the commercial environment and how representative the testing herds are of the target environments.

Genetic Parameters

Most estimates of heritability for production traits were within the accepted industry range (Ciobanu et al., 2011; Clutter, 2011), although this study was conducted under a strong health challenge. To the best of our knowledge, there are no estimates of genetic parameters for the novel resilience traits evaluated here in pigs. Variation for different traits has been explored as a potential indicator trait for resilience in dairy cattle. Green et al. (2004) evaluated the use of changes in somatic cell count (SCC) over time as an indicator for mastitis in lactating dairy cows and concluded that the maximum SCC and the standard deviation of log SCC were the best phenotypic indicators for incidence of mastitis. Recently, Elgersma et al. (2018) estimated genetic parameters for resilience traits from daily milk yield data from automated milking systems. Resilience indicators from milk yield data were calculated using the sum of “drop” days, negative slopes, and overall variation in milk yield calculated within lactation for each cow. Heritability estimates ranged from 0.06 to 0.10 and genetic correlations of variation in milk yield with udder health, ketosis, longevity, and persistency ranged from −0.29 to −0.52 (Elgersma et al., 2018). Elgersma et al. (2018), however, did not account for the individual cow milk yield trajectory over lactation when computing day-to-day variation in milk yield but targeted this for future research.

Heritability estimates for mortality and treatments in pigs are difficult to find in literature because of the swine pyramid structure, which results in most studies focusing on data collected in herds with limited disease. Guy et al. (2018) estimated the heritability of treatments to be between 0.04 and 0.06. Commercial test herds using the three-way terminal cross are becoming more popular in the swine industry but results from such data are not commonly reported in the literature. One example is Dufrasne et al. (2014), who used a sire model to estimate variance components for mortality (culling) traits. Heritability estimates ranged from 0.03 to 0.14 using threshold models (Dufrasne et al., 2014) but the rate of mortality after weaning was very low (<1%), which seems very unrealistic as typical commercial wean-to-finish barns have between 6 and 9% mortality on average (Stalder, 2017). Estimates of heritability for treatment and mortality could change with the amount of health challenge and incidence (Bishop and Woolliams, 2014). Companies will need to decide how much of a health challenge they need if they aim to select for resilience to disease. Challenging pigs too much comes at an economic and animal welfare cost. If not challenged enough, heritabilities of mortality, and treatments may become lower and response to selection will

be slowed (Mulder and Rashidi, 2017), although low heritabilities may be partially overcome with very large family sizes (many matings per sire). Treatment data is also challenging to collect in commercial testing systems. Many use mass treatments for disease outbreaks (e.g., feed and water medication). Water treatments may be more helpful for treatment under challenge, while feed medications may be more helpful for prevention (due to the off-feed events under challenge). Although it is possible to collect individual treatment data, commercial farms differ in the amount of data, and details they record. Factors such as withdrawal times may influence when and if an animal receives treatment. When animals are treated and/or euthanized is based in part on subjective decisions by farm or veterinary staff. If antibiotic free production is involved, this may also influence the decision to treat an animal or not. Resilience is expected to be more economically important in those conditions as management cannot mask the genetic potential for resilience.

Implementing Quantile Regression (QR) Phenotypes

One of the challenges when implementing QR phenotypes is that the quantile regression equation will depend on the severity of the disease challenge. For instance, if one barn is completely healthy over the years and another barn is severely challenged, the QR equation for each barn will be very different even if both are at the 5% level. This difference will be tied to how often contemporary groups are challenged and to what degree they are challenged. Mulder and Rashidi (2017) discussed the percentage of contemporary groups challenged and how that affects selection efficiency. When starting a commercial testing system, setting this initial QR threshold may be difficult. If the first groups are not heavily challenged, it will lead to setting the QR equation too high and capturing days that are not due to illness and other stressors, simply normal daily variation in FI. With a weaker disease challenge, a more appropriate QR may be 1%. One possibility is to create a training dataset for QR based on contemporary groups that were challenged and set the threshold based on that regression. Another possibility would be to take only healthy contemporary groups and set a lower bound threshold based on that data.

Heterogeneous Residual Variance in FI Data

One problem with using daily variation in FI vs. duration is that the variance of FI increases with age, which is not observed for daily duration data. This results in stressors having a greater impact on FI for older vs. younger pigs. As a result, $RMSE_{FI}$ puts a greater weight on later compared to earlier ages. Mean duration showed a slight negative trend with age and its variance was fairly constant across the finishing period. This may be one reason for the fairly low genetic correlation between FI and duration measures and could also explain why $RMSE_{DUR}$ had slightly higher genetic correlations with mortality and number of treatments than $RMSE_{FI}$, as $RMSE_{DUR}$ weights the early finishing period the same as the late finishing period. An attempt was made to adjust $RMSE_{FI}$ for this increasing variation over time, but this

still resulted in large outliers at later ages and did not improve the phenotype much in terms of genetic correlations to mortality and treatments (results not shown).

Causes of Mortality

Most recorded mortality reasons were linked to disease. Exceptions included death from blood sampling, rectal prolapse, fighting, and fracture/sprains, which amounted to ~11% of the 344 mortalities observed in these data. Removing mortality records due to non-health reasons, however, only resulted in small changes in estimates of heritability and genetic correlations, so they were left in the dataset for the current analyses, also because mortality by definition includes any pig that died regardless of cause. Although we typically think of mortality as health related, it is very multifactorial, as is sow mortality. One could decide to separate mortality by cause due to different genetic architecture for each cause and different economic weights (due to the average timing of death), but it is likely that the heritability would be even lower due to lower incidence, which would limit genetic selection. In addition, mortality for non-health related reasons could also have a genetic component. Treatments were almost exclusively linked to disease. Although some, such as lameness, could be argued to not be linked to disease, some diseases can be linked to lameness and removing them then becomes controversial.

Impact of Increased Variation in Performance

One major impact of disease is the increase in variation in production phenotypes such as growth, causing some pigs to be less than full value when harvested (Fix et al., 2010). In the current study, if pigs did not make weight, they were held in the finisher until they made the target weight range, resulting in more pigs achieving full value when harvested. The definition of full value is not consistent in research or the swine industry and was therefore avoided in this analysis. Some production systems require all animals to leave at a certain date regardless of weight (i.e., fixed time systems). This would lead to additional costs from disease as a result of greater variability in slaughter weights and carcass weights. Hubbs et al. (2008) used moments beyond the mean to include variance and skew for determining optimal marketing decisions and concluded these higher-order statistics appeared to be more important than they were in the past. Not only are carcasses lighter and therefore worth less in total, sort loss (or discount losses) from not meeting the optimal weight grid will also penalize these animals (Boys et al., 2007). Sometimes, these lightweight animals can go to alternative markets, but not always (Fix et al., 2010).

Use of Novel Traits in Healthy Nucleus Environments

The novel traits evaluated here can also be recorded in relatively healthy nucleus environments for stressors other than health challenges. A second major factor impacting these novel traits may be heat stress. The genetic correlation between RMSE in the nucleus vs. in a commercial environment (under disease challenge) will likely depend on the level and nature of the disease

challenge in the commercial environment and the amount of heat stress in each environment (among other stressors). Barns located in the Southern/SE USA region will be affected differently by heat stress than those in the upper Midwest or Canada. This barn was in Québec, Canada, and the heat stress experienced was minimal compared to other areas around the world. The novel resilience traits evaluated here will likely have lower means and be less variable and have lower heritability under nucleus conditions but will still include resilience to stressors.

CONCLUSIONS

Day-to-day variation (RMSE and QR) in feed intake or duration at the feeder can be used to quantify resilience in health challenged environments, such as a commercial testing scheme. The novel resilience phenotypes studied here were moderately heritable and genetically correlated with mortality and treatment rate. The genetic correlations reported here may underestimate true correlations because the initial challenge period was missed because pigs were first challenged in the nursery and feed intake data for RMSE and QR was recorded in the finishing unit only, while mortality and treatments were recorded over the entire wean-to-finish period. Many factors can cause variation in feed intake and in time at the feeder, including disease, heat stress, handling, and social interactions. Thus, the measures of resilience investigated here are still “black-box” phenotypes and should be viewed as general resilience instead of the narrower concept of disease resilience. Overall, daily variation in FI or associated duration data can be used to quantify resilience.

AUTHOR CONTRIBUTIONS

AP came up with the novel phenotypes, analyzed the data, and wrote the manuscript with help from JD. GP, MD, PGC, JH, FF, and JD designed the project and developed protocols for animal sourcing, management, and phenotype recording. JH was

in charge of veterinary oversight on the project. GP was in charge of the database and genotyping for the project. FF was the lead on day-to-day data collection and scheduling. All authors helped with interpretation of the results and reviewed and approved the final manuscript.

FUNDING

This project was funded by Genome Canada, Genome Alberta, Genome Prairie and PGC.

ACKNOWLEDGMENTS

The following individuals served as collaborators and representatives for member companies of the PigGen Canada Consortium. They participated in project and protocol development and implementation, coordinated the sources of piglets and collection of associated data, and contributed to the project through regular discussions during execution of the natural disease challenge project: Mr. D. Vandenbroek and Mr. B. DeVries, Alliance Genetics Canada, St. Thomas, ON, Canada; Dr. N. Dion and Ms. S. Blanchette, AlphaGene, St.-Hyacinthe, QU, Canada; Dr. T. Rathje, DNA Genetics, Columbus, NE, USA; Mr. M. Duggan, FastGenetics, Saskatoon, SK, Canada; Dr. R. Kemp, Genesus, London, ON, Canada; Dr. P. Charagu, Hypor, Regina, SK, Canada; and Dr. P. Mathur, Topigs Norsvin, Helvoirt, The Netherlands. Mr. Michael Lowings at Delta Genomics (Edmonton, AB, Canada) is acknowledged for processing the raw genotypes from the Affymetrix SNP chip and Mr. Patrick Gagnon at CDPQ for processing the raw feed intake data. Thanks to Mr. Jason Grant at the University of Alberta for managing the database for the project and helping with genotyping and SNP map file questions. Special thanks to Dr. Wendy Rauw for helpful discussions on resilience and from her presentation on heat tolerant vs. non-heat tolerant growing pigs, which inspired the RMSE measure used in the present study.

REFERENCES

- Albers, G. A., Gray, G. D., Piper, L. R., Barker, J. S. F., Le Jambre, L. F., and Barger, I. A. (1987). The genetics of resistance and resilience to *Haemonchus contortus* infection in young merino sheep. *Int. J. Parasito.* 17, 1355–1363. doi: 10.1016/0020-7519(87)90103-2
- Bishop, S. (2012). A consideration of resistance and tolerance for ruminant nematode infections. *Front. Genet.* 3:168. doi: 10.3389/fgene.2012.00168
- Bishop, S., and Woolliams, J. A. (2014). Genomics and disease resistance studies in livestock. *Livest. Sci.* 166, 190–198. doi: 10.1016/j.livsci.2014.04.034
- Bouwman, A. C., Bergsma, R., Duijvesteijn, N., and Bijma, P. (2010). Maternal and social genetic effects on average daily gain of piglets from birth until weaning. *J. Anim. Sci.* 88, 2883–2892. doi: 10.2527/jas.2009-2494
- Boys, K., Li, N., Preckel, P. V., Schinckel, A. P., and Foster, K. A. (2007). Economic replacement of a heterogeneous herd. *Am. J. of Agr. Econ.* 89, 24–35. doi: 10.1111/j.1467-8276.2007.00960.x
- Cade, B. S., and Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* 1, 412–420. doi: 10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2
- Cai, W., Casey, D. S., and Dekkers, J. C. M. (2008). Selection response and genetic parameters for residual feed intake in Yorkshire swine. *J. Anim. Sci.* 86, 287–298. doi: 10.2527/jas.2007-0396
- Casey, D. S., Stern, H. S., and Dekkers, J. C. M. (2005). Identification of errors and factors associated with errors in data from electronic swine feeders. *J. Anim. Sci.* 83, 969–982. doi: 10.2527/2005.835969x
- Christensen, O. F., and Lund, M. S. (2010). Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2. doi: 10.1186/1297-9686-42-2
- Ciobanu, D. C., Lonergan, S. M., and Huff-Lonergan, E. J. (2011). “Genetics of meat quality and carcass traits,” in *The Genetics of the Pig 2nd Edn*, eds M. F. Rothschild and A. Ruvinsky (Oxfordshire: CAB International), 355–389.
- Clutter, A. C. (2011). “Genetics of performance traits,” in *The Genetics of the Pig 2nd Edn*, eds M. F. Rothschild and A. Ruvinsky (Oxfordshire: CAB International), 325–354.
- Colditz, I. G., and Hine, B. C. (2016). Resilience in farm animals: biology, management, breeding and implications for animal welfare. *An. Prod. Sci.* 56, 1961–1983. doi: 10.1071/AN15297
- de Haer, L. C. M., Luiting, P., and Aarts, H. L. M. (1993). Relations among individual (residual) feed intake, growth performance and feed intake pattern of growing pigs in group housing. *Livest. Prod. Sci.* 36, 233–253. doi: 10.1016/0301-6226(93)90056-N
- Doeschl-Wilson, A. B., Villanueva, B., and Kyriazakis, I. (2012). The first step toward genetic selection for host tolerance to infectious pathogens: obtaining the tolerance phenotype through group estimates. *Front. Genet.* 3:265. doi: 10.3389/fgene.2012.00265

- Dufresne, M., Misztal, I., Tsuruta, S., Gengler, N., and Gray, K. A. (2014). Genetic analysis of pig survival up to commercial weight in a crossbred population. *Livest. Sci.* 167, 19–24. doi: 10.1016/j.livsci.2014.05.001
- Elgersma, G. G., de Jong, G., van der Linde, R., and Mulder, H. A. (2018). Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. *J. Dairy Sci.* 101, 1240–1250. doi: 10.3168/jds.2017-13270
- Fix, J. S., Cassady, J. P., Holl, J. W., Herring, W. O., Culbertson, M. S., and See, M. T. (2010). Effect of piglet birth weight on survival and quality of commercial market swine. *Livest. Sci.* 132, 98–106. doi: 10.1016/j.livsci.2010.05.007
- Fragomeni, B. O., Lourenco, D. A. L., Tsuruta, S., Andonov, S., Gray, K., Huang, Y., et al. (2016). Modeling response to heat stress in pigs from nucleus and commercial farms in different locations in the United States. *J. Anim. Sci.* 94, 4789–4798. doi: 10.2527/jas.2016-0536
- Frisch, J. E. (1981). Changes occurring in cattle as a consequence of selection for growth rate in a stressful environment. *J. Agri. Sci.* 91, 23–38. doi: 10.1017/S0021859600031841
- Gilbert, H., Billon, Y., Brossard, L., Faure, J., Gatellier, P., Gondret, F., et al. (2017). Review: divergent selection for residual feed intake in the growing pig. *Animal* 11:9. doi: 10.1017/S175173111600286X
- Green, M. J., Green, L. E., Schukken, Y. H., Bradley, A. J., Peeler, E. J., Barkema, H. W., et al. (2004). Somatic cell count distributions during lactation predict clinical mastitis. *J. Dairy. Sci.* 87, 1256–1264. doi: 10.3168/jds.S0022-0302(04)73276-2
- Guy, S. Z. Y., Li, L., Thomson, P. C., and Hermes, S. (2017). Contemporary group estimates adjusted for climatic effects provide a finer definition of the unknown environmental challenges experienced by growing pigs. *J. Anim. Breed. Genet.* 134, 520–530. doi: 10.1111/jbg.12282
- Guy, S. Z. Y., Li, L., Thomson, P. C., and Hermes, S. (2018). “Genetic parameters for health of the growing pig using medication records,” in *Proceeding 11th World Congress of Genetics Applied to Livestock Production*, (Auckland).
- Guy, S. Z. Y., Thomson, P. C., and Hermes, S. (2012). Selection of pigs for improved coping with health and environmental challenges: breeding for resistance or tolerance? *Front. Genet.* 3:281. doi: 10.3389/fgene.2012.00281
- Hemsworth, P. H., Rice, M., Nash, J., Giri, K., Butler, K. L., A. J., et al. (2013). Effects of group size and floor space allowance on grouped sows: aggression, stress, skin injuries, and reproductive performance. *J. Anim. Sci.* 91, 4953–4964. doi: 10.2527/jas.2012-5807
- Holtkamp, D. J., Kliebenstein, J. B., Neumann, E. J., Zimmerman, J. J., Rotto, H. F., Yoder, T. K., et al. (2013). Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J. Swine Health Prod.* 21, 72–84. Available online at: <https://www.aasv.org/shap/issues/v21n2/v21n2p72.html>
- Hubbs, T., Preckel, P. V., Schinckel, A. P., Deen, J., Foster, K. A., Curtis, S. E., et al. (2008). “The new economics of livestock production management,” in *American Agricultural Economics Association Annual Meeting*, (Orlando, FL).
- Hyun, Y., Ellis, M., Riskowski, G., and Johnson, R. W. (1998). Growth performance of pigs subjected to multiple concurrent environmental stressors. *J. Anim. Sci.* 76, 721–727. doi: 10.2527/1998.763721x
- King, M. T. M., LeBlanc, S. J., Pajor, E. A., Wright, T. C., and DeVries, T. J. (2018). Behavior and productivity of cows milked in automated systems before diagnosis of health disorders in early lactation. *J. Dairy. Sci.* 101, 1–14. doi: 10.3168/jds.2017-13686
- Knap, P. W. (2009). “Voluntary feed intake in pig breeding,” in *Voluntary Feed Intake in Pigs*, ed. D. Torrallardona and E. Roura (Wageningen: Wageningen Academic Publishers), 13–35.
- Kyriazakis, I., and Doeschl-Wilson, A. (2009). “Anorexia during infection in mammals: variation and its sources,” in *Voluntary Feed Intake in Pigs*, ed. D. Torrallardona and E. Roura (Wageningen: Wageningen Academic Publishers), 307–321.
- Kyriazakis, I., and Tolamp, B. (2018). “Hunger and Thirst,” in *Animal Welfare*, ed. M. C. Appleby, I. A. S. Olsson, and F. Galindo. (Oxfordshire: CAP International), 56–75.
- Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663. doi: 10.3168/jds.2009-2061
- Lough, G., Kyriazakis, I., Bergmann, S., Lengeling, A., and Doeschl-Wilson, A. B. (2015). Health trajectories reveal the dynamic contributions of host genetic resistance and tolerance to infection outcome. *Proc. R. Soc. B.* 282:20152151. doi: 10.1098/rspb.2015.2151
- Lu, D., Jiao, S., Tiezzi, F., Knauer, M., Huang, Y., Gray, K. A., et al. (2017). The relationship between different measures of feed efficiency and feed behavior traits in Duroc pigs. *J. Anim. Sci.* 95, 3370–3380. doi: 10.2527/jas2017.1509
- Madsen, T. N., and Kristensen, A. R. (2005). A model for monitoring the condition of young pigs by their drinking behaviour. *Comp. Electr. Agricult.* 48, 138–154. doi: 10.1016/j.compag.2005.02.014
- Martínez-Miró, S., Tecles, F., Ramón, M., Escribano, D., Hernández, F., Madrid, J., et al. (2016). Causes, consequences, and biomarkers of stress in swine: an update. *BMC Vet. Res.* 12:171. doi: 10.1186/s12917-016-0791-8
- Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., and Lee, D. H. (2002). “BLUPF90 and related programs (BGF90),” in *Proceeding 7th World Congress of Genetics Applied to Livestock Production*, (Montpellier).
- Mulder, H. A., and Rashidi, H. (2017). Selection on resilience improves disease resistance and tolerance to infections. *J. Anim. Sci.* 95, 3346–3358. doi: 10.2527/jas2017.1479
- Ødegård, J., Meuwissen, T. H. E., and Heringstad, B., Madsen, P. (2010). A simple algorithm to estimate genetic variance in an animal threshold model using Bayesian inference. *Genet. Sel. Evol.* 42:29. doi: 10.1186/1297-9686-42-29
- Petry, A., McGilvray, W., Rakhshandeh, A. R., and Rakhshandeh, A. (2017). Technical note: assessment of an alternative technique for measuring body temperature in pigs. *J. Anim. Sci.* 95, 3270–3274. doi: 10.2527/jas2017.1566
- Petry, D. B., Holl, J. W., Weber, J. S., Doster, A. R., Osorio, F. A., and Johnson, R. K. (2005). Biological responses to porcine respiratory and reproductive syndrome virus in pigs of two genetic populations. *J. Anim. Sci.* 83, 1494–1502. doi: 10.2527/2005.8371494x
- Petry, D. B., Lunney, J., Boyd, P., Kuhar, D., Blankenship, E., and Johnson, R. K. (2007). Differential immunity in pigs with high and low responses to porcine reproductive and respiratory syndrome virus infection. *J. Anim. Sci.* 85, 2075–2092. doi: 10.2527/jas.2006-721
- Pomar, C., and Marcoux, M. (2003). Comparing the Canadian pork lean yields and grading indexes predicted from grading methods based on Destron and Hennessey probe measurements. *Can. J. Anim. Sci.* 83, 451–458. doi: 10.4141/A02-107
- R core team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available online at: <https://www.R-project.org/>
- Rauw, W. M., Johnson, A. K., Gomez-Raya, L., and Dekkers, J. C. M. (2017a). A hypothesis and review of the relationship between selection for improved production efficiency, copying behavior, and domestication. *Front. Genet.* 8:134. doi: 10.3389/fgene.2017.00134
- Rauw, W. M., Mayorga, E. J., Lei, S. M., Dekkers, J. C. M., Patience, J. F., Gabler, N. K., et al. (2017b). Effects of diet and genetics on growth performance of pigs in response to repeated exposure to heat stress. *Front. Genet.* 8:155. doi: 10.3389/fgene.2017.00155
- Rusakovica, J., Kremer, V. D., Plotz, T., Rohlf, P., and Kyriazakis, I. (2017). The genetic basis of novel water utilization and drinking behaviour traits and their relationship with biological performance in turkeys. *Genet. Sel. Evol.* 49:72. doi: 10.1186/s12711-017-0343-0
- Sadler, L. J., Johnson, A. K., Lonergan, S. M., Nettleton, D., and Dekkers, J. C. M. (2011). The effect of selection for residual feed intake on general behavioral activity and the occurrence of lesions in Yorkshire gilts. *J. Anim. Sci.* 89, 258–266. doi: 10.2527/jas.2009-2595
- Salak-Johnson, J. L., and McGlone, J. J. (2007). Making sense of apparently conflicting data: stress and immunity in swine and cattle. *J. Anim. Sci.* 85(E. Suppl.), E81–E88. doi: 10.2527/jas.2006-538
- Sandberg, F. B., Emmans, G. C., and Kyriazakis, I. (2006). A model for predicting feed intake of growing animals during exposure to pathogens. *J. Anim. Sci.* 84, 1552–1566. doi: 10.2527/2006.8461552x
- Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genom.* 15:478. doi: 10.1186/1471-2164-15-478
- Schulze, V., Rohe, R., Lorenzo Bermejo, J., Looft, H., and Kalm, E. (2003). The influence of feeding behavior on feed intake curve parameters and performance traits of station-tested boars. *Livest. Prod. Sci.* 82, 105–116. doi: 10.1016/S0301-6226(03)00034-4

- Smith, T. K., McMillan, E. G., and Castillo, J. B. (1997). Effect of feeding blends of *Fusarium* mycotoxin-contaminated grains containing deoxynivalenol and fusaric acid on growth and feed consumption of immature swine. *J. Anim. Sci.* 75, 184–2191. doi: 10.2527/1997.7582184x
- Stalder, K. J. (2017). *Pork Industry Productivity Analysis*. National Pork Board. Available online at: <https://www.pork.org/wp-content/uploads/2017/11/2017-pork-industry-productivity-analysis.pdf>
- St-Pierre, N. R., Cobanov, B., and Schnitkey, G. (2003). Economic losses from heat stress by US livestock industries. *J. Dairy. Sci.* 86(E. Suppl), E52–E77. doi: 10.3168/jds.S0022-0302(03)74040-5
- Street, B. R., and Gonyou, H. W. (2008). Effects of housing finishing pigs in two group sizes and at two floor space allocations on production, health, behavior, and physiological variables. *J. Anim. Sci.* 86, 982–992. doi: 10.2527/jas.2007-0449
- van der Waaij, E. H., Bijma, P., Bishop, S. C., and van Arendonk, J. A. M. (2000). Modeling selection for production traits under constant infection pressure. *J. Anim. Sci.* 78, 2809–2820. doi: 10.2527/2000.78112809x
- VanRaden, P. (2008). Efficient methods to compute genomic prediction. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Vilas Boas Ribeiro, B. P., Lanferdini, E., Palencia, J. Y. P., Lemes, M. A. G., Teixeira de Abreu, M. L., de Souza Cantarelli, V., et al. (2018). Heat negatively affects lactating swine: A meta-analysis. *J. Therm. Biol.* 74, 325–330. doi: 10.1016/j.jtherbio.2018.04.015
- Von Felde, A., Roehe, R., Looft, H., and Kalm, E. (1996). Genetic association between feed intake and feed intake behavior at different stages of growth of group-housed boars. *Livest. Prod. Sci.* 47, 11–22. doi: 10.1016/S0301-6226(96)01006-8
- Webel, D. M., Finck, B. N., Baker, D. H., and Johnson, R. W. (1997). Time course of increased plasma cytokines, cortisol, and urea nitrogen in pigs following intraperitoneal injection of lipopolysaccharide. *J. Anim. Sci.* 75, 1514–1520. doi: 10.2527/1997.7561514x
- Young, J. M., Cai, W., and Dekkers, J. C. M. (2011). Effect of selection for residual feed intake on feeding behaviour and daily feed intake patterns in Yorkshire swine. *J. Anim. Sci.* 89, 639–647. doi: 10.2527/jas.2010-2892
- Zimmerman, J. J., Karriker, L. A., Ramirez, A., Schwartz, K. J., and Stevenson, G. W. (2012). *Diseases of Swine, 10th edn*. West Sussex: Wiley-Blackwell

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past co-authorship with one of the authors JD.

Copyright © 2019 Putz, Harding, Dyck, Fortin, Plastow, Dekkers and PigGen Canada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Opportunities to Improve Resilience in Animal Breeding Programs

Tom V. L. Berghof*, Marieke Poppe and Han A. Mulder

Wageningen University & Research Animal Breeding and Genomics, Wageningen, Netherlands

OPEN ACCESS

Edited by:

Andrea B. Doeschl-Wilson,
Roslin Institute, University of
Edinburgh, United Kingdom

Reviewed by:

Jiuzhou Song,
University of Maryland, College Park,
United States
Jack Dekkers,
Iowa State University, United States
Pieter Willem Knap,
Genus-PIC, Germany

*Correspondence:

Tom V. L. Berghof
tom.berghof@wur.nl

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 26 June 2018

Accepted: 11 December 2018

Published: 14 January 2019

Citation:

Berghof TVL, Poppe M and
Mulder HA (2019) Opportunities to
Improve Resilience in Animal Breeding
Programs. *Front. Genet.* 9:692.
doi: 10.3389/fgene.2018.00692

Resilience is the capacity of an animal to be minimally affected by disturbances or to rapidly return to the state pertained before exposure to a disturbance. However, indicators for general resilience to environmental disturbances have not yet been defined, and perhaps therefore resilience is not yet included in breeding goals. The current developments on big data collection give opportunities to determine new resilience indicators based on longitudinal data, which can aid to incorporate resilience in animal breeding goals. The objectives of this paper were: (1) to define resilience indicator traits based on big data, (2) to define economic values for resilience, and (3) to show the potential to improve resilience of livestock through inclusion of resilience in breeding goals. Resilience might be measured based on deviations from expected production levels over a period of time. Suitable resilience indicators could be the variance of deviations, the autocorrelation of deviations, the skewness of deviations, and the slope of a reaction norm. These (new) resilience indicators provide opportunity to include resilience in breeding programs. Economic values of resilience indicators in the selection index can be calculated based on reduced costs due to labor and treatments. For example, when labor time is restricted, the economic value of resilience increases with an increasing number of animals per farm, and can become as large as the economic value of production. This shows the importance of including resilience in breeding goals. Two scenarios were described to show the additional benefit of including resilience in breeding programs. These examples showed that it is hard to improve resilience with only production traits in the selection index, but that it is possible to greatly improve resilience by including resilience indicators in the selection index. However, when health-related traits are already present in the selection index, the effect is smaller. Nevertheless, inclusion of resilience indicators in the selection index increases the response in the breeding goal and resilience, which results in less labor-demanding, and thus easier-to-manage livestock. Current developments on massive collection of data, and new phenotypes based on these data, offer exciting opportunities to breed for improved resilience of livestock.

Keywords: resilience, livestock, breeding program, micro-environment, macro-environment, economic value, big data, longitudinal data

INTRODUCTION

Modern livestock production is characterized by intensification, i.e., a higher number of animals per farm. To achieve successful intensification of livestock production, without negative effects on animals, farmers and farms, certain requirements need to be met. One of these requirements for intensification of livestock production is the capability of the farmer to take care of a larger number of animals. This requires healthy and easy-to-manage animals that need little/less attention time (Elgersma et al., 2018). Resilient animals are animals that need little/less attention time: increasing resilience is therefore desired. Improvement of resilience can be realized by different strategies. One strategy is to increase resilience by genetic selection in breeding programs. The advantage of genetic selection, in contrast to management improvements, is that it is cumulative and affects all subsequent generations of livestock.

We define resilience as “the capacity of the animal to be minimally affected by disturbances or to rapidly return to the state pertained before exposure to a disturbance” (adjusted from Colditz and Hine, 2016). Several definitions of resilience (e.g., Colditz and Hine, 2016), and resilience-associated concepts have been discussed in literature: robustness, tolerance, resistance, G×E interaction, genetic heterogeneity of environmental variance, plasticity, environmental sensitivity, canalization, (developmental) stability, and residual within-individual phenotypic variation (e.g., Holling, 1973; Debat and David, 2001; De Jong and Bijma, 2002; Flatt, 2005; Knap, 2005; Mulder et al., 2007, 2013; Bishop, 2012; Westneat et al., 2015; Colditz and Hine, 2016; Marjanovic et al., 2018). We have summarized these definitions and concepts in **Box 1**, but it is beyond the scope of this paper to discuss differences and similarities between these.

Disturbances can be of different nature, being either physical (e.g., disease, temperature stress) or psychological (e.g., novel environment, social stressor, human interaction) (see Colditz and Hine, 2016 for a review). “General” resilience is therefore a composite trait, consisting of different resilience types depending on the nature of the disturbance (Colditz and Hine, 2016; Elgersma et al., 2018). Disturbances can be categorized as “macro-environmental factors” or “micro-environmental factors” (Falconer and Mackay, 1996; Mulder et al., 2013). Macro-environmental factors are characteristics of an environment and thus affect the (majority of the) whole population within that macro-environment (e.g., disease pressure, ambient temperature). Genetic variation in the response to these macro-environmental factors can be expressed as the genetic variance of the slope of a reaction norm over different environments or different quantities of a disturbance (Mulder et al., 2013). Micro-environmental factors occur within a macro-environment, and thus affect only a minority of the whole population within that macro-environment (e.g., diseases, social interactions). Genetic differences in response to micro-environmental factors can be expressed as genetic variance in the size of environmental variance (Mulder et al., 2013). Despite the fact that macro- and micro-environmental sensitivities refer to different concepts and are best investigated with different

methods (Mulder et al., 2015a), the estimated genetic correlation between them was high (0.76 with SE = 0.10; Mulder et al., 2013). This indicates that, even though disturbances have effects on a different scale, resilience to these disturbances have a common genetic background and will respond to selection for increased resilience in a similar direction. Nevertheless, from a practical point of view, resilience to occasional macro-environmental disturbances, such as disease outbreaks and heat waves, are less frequent and therefore of lesser importance, at least in most farms in temperate climates. Thus easy-to-manage livestock is livestock with increased resilience to day-to-day micro-environmental disturbances within the macro-environment (i.e., a farm), but also with increased resilience to macro-environmental disturbances.

Resilience indicators are not yet included in selection indices as far as we know, despite their clear relevance for healthy and easy-to-manage livestock (and also production uniformity and production efficiency). This is likely due to ignorance on how to define, measure and weight resilience indicators and their economic values, and due to the belief that current health-related traits in selection indices (e.g., longevity, mortality, growth) cover resilience, at least mostly (e.g., Knap, 2005). Although suggestions for measuring environmental sensitivity (Knap, 2009b) (i.e., resilience) and determining its economic value have been proposed (Knap, 2005), proper data collection and data processing tools were considered to be major challenges for implementation (Knap, 2009b). However, recent technological developments allow collection of big/longitudinal data and derivation of new phenotypes from it (Mulder, 2017). These developments will only continue to expand and will become more important in the future. New ideas to explore and exploit big data and new phenotypes are required. Also, health-related traits do not necessarily reflect general resilience, but mainly capture resistance to diseases (i.e., disease resilience; Mulder and Rashidi (2017). In addition, breeding value estimation for disease-related traits is limited to only a few diseases, if any. Thus there seems to be additional benefit for including general resilience indicators in selection indices.

The objectives of this paper were therefore:

1. to define resilience indicator traits based on big/longitudinal data,
2. to define economic values for resilience indicators based on labor and treatment costs, and
3. to show the potential to improve resilience of livestock through inclusion of resilience in breeding goals.

The paper addresses the described objectives sequentially and ends with overall conclusions, and therefore the paper differs in its setup from conventional research papers.

DEFINING RESILIENCE INDICATORS

Resilience has been a popular concept in a broad range of scientific disciplines in the last two decades (e.g., Scheffer et al., 2015; Ge et al., 2016). However, defining resilience indicators has been proven to be difficult, and different options have been proposed, which suggests that the “right” resilience indicator

BOX 1 | Definitions of resilience and resilience-associated concepts discussed in literature (quoted). Note that this box does not serve as a critical reflection of literature, but as an overview for interpretation of different concepts.

Resilience: the capacity of the animal to be minimally affected by disturbances or to rapidly return to the state pertained before exposure to a disturbance (adjusted from Colditz and Hine, 2016; this paper).

Robustness: animals that combine high production potential with resilience to external stressors, allowing for unproblematic expression of high production potential in a wide variety of environmental conditions (adjusted from Knap, 2005).

Tolerance: the net impact on performance of a given level of disturbance (adjusted from Bishop, 2012).

Resistance: the ability of the host animal to exert control over the disturbance (adjusted from Bishop, 2012).

GxE interaction: the best genotype in one environment may not be the best genotype in another environment and those genotypes differ in their response to environmental factors (Mulder et al., 2013).

Environmental sensitivity: loss of flexibility to deal with intensive or limiting conditions, due to unbalanced resource allocation (Knap, 2005).

Heritable variation in environmental variance or inherited variability: the variability of trait values of a genotype, measured either repeatedly on the same individual, or on multiple individuals belonging to the same family (Marjanovic et al., 2018).

(Phenotypic) plasticity: the ability of an organism to alter its physiology, morphology or development in response to changes in its environment (Debat and David, 2001).

Macro-environmental sensitivity: the property of the organisms of a genotype to develop systematically different phenotypes in different environments = norm of reaction (Debat and David, 2001).

(Genetic) canalization: similarity of the expression of the phenotypic character under different conditions of development; the process by which phenotypic variation is reduced by developmental mechanisms or a genotype's phenotype remains relatively invariant when individuals of the same single- or multilocus genotype differ in their genetic background; a particular kind of epistasis (Debat and David, 2001; Flatt, 2005).

(Developmental) stability: the ability of a system to return to an equilibrium state after a temporary disturbance (Holling, 1973).

Residual within-individual phenotypic variation: amount of within-individual variance not explained in a specific statistical model (i.e., the average squared deviations of observations from an individual's reaction norm), averaged over a sample of individuals (Westneat et al., 2015).

has yet to be defined. Recent technological developments give new opportunities to explore alternative resilience indicators based on longitudinal data (Mulder, 2017). Here we propose resilience indicators, which exploit the availability of many repeated observations per individual or per (sire-)family (i.e., big data). We will also describe some of the conditions these resilience indicators will have to fulfill in order to be informative, and their potential weaknesses.

Resilience Indicators

Measuring general resilience is difficult. Many studies have focused on one particular type of resilience (especially disease resilience) and have used experimental set-ups to identify underlying physiological mechanisms. However, these mechanisms strongly depend on the nature of the disturbance, are often chosen based on the interest of the study, and might characterize a phenotype that is (too) dependent on the investigated disturbance (Colditz and Hine, 2016). Even though these studies can provide useful information and insights in physiology, results may not be representative of general resilience. Furthermore, typically these challenge environments deviate from commercial environments and might therefore be less representative due to GxE interactions. Instead it has been proposed to measure “*summary characteristics of response variables*” (quoted from Colditz and Hine, 2016) as general resilience indicators (Knap, 2009b; Doeschl-Wilson et al., 2012; Colditz and Hine, 2016; Elgersma et al., 2018).

Colditz and Hine (2016) proposed a diverse set of “*summary characteristics of response variables*” to measure resilience to a disturbance, including typical production traits like feed intake, growth rate, and other production variables. Individuals

experiencing a disturbance eat and produce less than (i.e., they deviate from) their potential without a disturbance (e.g., Van der Waaij et al., 2000). However, a deviation between observed production mean and estimated potential (as proposed by Colditz and Hine, 2016) does not necessarily fully cover the definition of resilience: resilient animals might have a severe drop in production, but might also have the capacity to rapidly return to the state pertained before exposure to a disturbance compared to less resilient animals. Deviations over a period of time (for at least the length of the disturbance) likely reflect resilience better. Thus, resilience might be measured based on deviations of expected production and observed production (i.e., residuals) over a period of time.

Many potential resilience indicator traits have been described, aimed at predicting changes in states of ecosystems (see Table 1 of Scheffer et al., 2015 for an elaborate overview). However, many of these resilience indicators are difficult to apply to livestock species or genetic improvement of livestock, because the resilience indicators are not suitable in livestock production data (e.g., spatial correlation, spectral reddening). In addition, a large number of investigated resilience indicators are alike, because they use the data in a similar way, i.e., the number of unique indicators is limited. However, we do see opportunities for some of them. In this paper, we elaborated on four resilience indicators, for which we propose they show a different perspective of resilience [see also Scheffer et al. (2018)]. Suitable resilience indicators based on a single trait (e.g., production, feed intake) in animal production might be *variance* of deviations, *autocorrelation* of deviations, *skewness* of deviations of production traits over a period of time, and the *slope* of a reaction norm (see Table 1). We assume that in general

TABLE 1 | Overview of possible resilience indicators based on big/longitudinal data per animal with description, interpretation and remarks.

| Indicators | Description | Interpretation | Remarks |
|---|---|---|---|
| Variance of deviations | Indication of the impact of disturbances | Low(er) variance for animals without disturbances or not influenced by disturbances; High(er) variance for animals influenced by disturbances; Captures the severity and duration (i.e., rate of recovery) of environmental perturbations an individual experiences. | Does not discriminate between negative and positive deviations; Also known as uniformity, inherited variability, residual variance, or environmental variance. |
| (Lag-one) Autocorrelation of deviations | Indication of the length of the impact of disturbances | An autocorrelation around 0 (i.e., subsequent deviations are unrelated) for animals without disturbances, not influenced by disturbances, or with a fast recovery from disturbances; An autocorrelation toward +1 (i.e., subsequent deviations/observations are more alike) for animals influenced by disturbances and with slow(er) recovery from disturbances; An autocorrelation toward -1 (i.e., subsequent deviations are opposite) for animals influenced by disturbances and a fast and over-compensating response to disturbances, e.g., Compensatory growth; Captures the duration (i.e., rate of recovery) of environmental perturbations an individual experiences. | Does not discriminate between negative and positive deviations; Does not give an indication of the impact of deviations. |
| Skewness of deviations | Indication of the direction of disturbances | A skewness around 0 for animals without disturbances or not influenced by disturbances; A positive skewness for animals having (mainly) positive deviations due to positive responses to environmental improvements; A negative skewness for animals having (mainly) negative deviations due to disturbances; Captures the severity of environmental perturbations an individual experiences. | Might be more sensitive to outliers. |
| Slope of a reaction norm | Indication of resilience toward a macro-environmental disturbance | A slope of 0 for animals not influenced by the disturbance; A slope below 0 for animals influenced by the disturbance with steeper, negative slopes for animals more influenced; Captures the severity of a macro-environmental perturbations an individual experiences. | Detection of disturbance can only be done on farm level; Each disturbance has its own slope, i.e., multiple slopes per individual are possible; Disentanglement of two disturbances simultaneously is not possible; Micro-environmental disturbances are not captured by the slope of a reaction norm, i.e., no general resilience indication. |

We assume that in general negative deviations are more observed for less resilient animals, but this might be opposite depending on the observed trait.

negative deviations are mostly observed for less resilient animals, but this might be opposite depending on the trait, e.g., body temperature.

- **Variance** of deviations (also known as uniformity, inherited variability, residual variance, or genetic heterogeneity of environmental variance (see Hill and Mulder, 2010; Elgersma et al., 2018 for overviews) gives an indication of the impact of disturbances with: low(er) variance for animals without disturbances or not influenced by disturbances; and high(er) variance for animals influenced by disturbances. We hypothesize that the variance of deviations captures the severity and duration (i.e., rate of recovery) of environmental perturbations an individual experiences. The contribution of severity and duration cannot be disentangled within this resilience indicator.
- (Lag-one) **autocorrelation** of deviations gives an indication of the length of the impact of disturbances with: an autocorrelation around 0 (i.e., subsequent deviations are unrelated) for animals without disturbances, not influenced by disturbances, or with a fast recovery from disturbances;

an autocorrelation toward +1 (i.e., subsequent deviations are more alike) for animals influenced by disturbances and with slow(er) recovery from disturbances; and an autocorrelation toward -1 (i.e., subsequent deviations are opposite) for animals influenced by disturbances and a fast and over-compensating response to disturbances, e.g., compensatory growth. We hypothesize that the autocorrelation of deviations captures the duration (i.e., rate of recovery) of environmental perturbations an individual experiences.

- **Skewness** of deviations gives an indication of the direction of deviations with: a skewness around 0 for animals without disturbances or not influenced by disturbances; a positive skewness for animals having (mainly) positive deviations due to positive responses to environmental improvements; and a negative skewness for animals having (mainly) negative deviations due to disturbances. We hypothesize that the skewness of deviations captures the severity of environmental perturbations an individual experiences.
- **Slope** of a reaction norm gives an indication of resilience toward a macro-environmental disturbance

(Finlay and Wilkinson, 1963), e.g., disease pressure (Herrero-Medrano et al., 2015) or a heat wave (Ravagnolo and Misztal, 2000). The slope of a reaction norm is estimated based on the production of an individual given the level of a disturbance with: a slope of 0 for animals not influenced by the disturbance, and a slope below 0 for animals influenced by the disturbance with steeper, negative slopes for animals that are influenced more. We hypothesize that the slope of a reaction norm captures the severity of a macro-environmental perturbations an individual experiences.

Thus, less resilient animals are expected to have a larger variance, a positive autocorrelation, a negative skewness, and a steeper slope than the population average, assuming that the disturbance is reducing the trait value. Resilient animals are expected to have a smaller variance (i.e., closer to 0), an autocorrelation and a skewness around 0, and a slope closer to zero than the population average.

Some remarks regarding the proposed resilience indicators need to be made. A general remark to be made is that scaling can cause that animals with a higher mean have a higher variance of deviations, but they may also be genetically more sensitive, less resilient compared to animals with a low production (Falconer and Mackay, 1996). Specific remarks to be made are: the underlying assumption for both variance and autocorrelation as resilience indicators is that deviations are mainly (or solely) negative, because disturbances will reduce production. However, variance and autocorrelation do not discriminate between negative and positive deviations. Therefore, animals showing production above the expected production, like compensatory production (perhaps a desired trait, part of resilience), will be characterized as not/less resilient. On the other hand, such animals are likely to have shown a severe reduction in production due to the disturbance, clearly showing lack of resistance against the disturbance. In contrast, skewness does differentiate between positive and negative deviations, but is likely to be more sensitive to outliers due to cubic terms. The autocorrelation can also be negative, meaning that an animal shows fast and over-compensating responses to disturbances, e.g., compensatory growth. Although a negative autocorrelation is preferred over a (strong) positive autocorrelation, an animal with an autocorrelation around 0 is not affected by a disturbance at all, and is therefore preferred over a negative autocorrelation. Finally, the estimation of the slope of a reaction norm requires quantification on farm level either by a known environmental covariate, e.g., temperature (Ravagnolo and Misztal, 2000; Bohmanova et al., 2007; Carabaño et al., 2017), or by an overall drop in production on a farm as a consequence of many individual drops in production on that farm (Rashidi et al., 2014; Herrero-Medrano et al., 2015). In addition, each disturbance has its own slope, and it is also almost impossible to estimate multiple slopes for multiple disturbances occurring at the same time. However, the slope of a reaction norm can also be interpreted as (a form of) general resilience, although the slope likely does not capture the response to micro-environmental disturbances, which occur at the individual animal level, e.g., endemic diseases. Thus the slope only provides information for disturbances

that cause a decline in farm performance, such as heat stress or a disease outbreak (in case of e.g., heat stress). Thereby the slope provides only information on macro-environmental disturbances. This is in contrast to the other three proposed resilience indicators based on deviations in **Table 1**, that can be used for epidemic (at the farm level) and endemic events (at the individual level). Therefore, these indicators provide information for estimation of breeding values on all types of disturbances and thus general resilience. But perhaps the “best” resilience indicator is yet to be defined based on a combination of the different resilience indicators: a multivariate approach of resilience, for example based on cross-correlations or eigenvalues which both capture the relation(s) between different resilience indicators based on different phenotypes (e.g., Lade and Gross, 2012; Scheffer et al., 2015; Gijzel et al., 2017), or an index of a number of indicators (i.e., a selection index approach, Hazel, 1943).

Little is known about the usefulness of the proposed resilience indicators based on deviations for livestock genetics: no studies have investigated the autocorrelations of deviations and the skewness of deviations, and only a few studies have investigated the variance of deviations. Elgersma et al. (2018) investigated the raw phenotypic variance of daily milk yield over the whole milking period and its relation to health and longevity traits in dairy cows. The raw variance of milk yield was heritable (0.10; Elgersma et al., 2018). Moreover, on a genetic level, cows with a low variance in milk yield deviations had significant fewer production-related diseases [i.e., udder health, genetic correlation (r_g) = -0.36 ; ketosis, r_g = -0.52] and a higher longevity (r_g = -0.30), suggesting a higher resilience (Elgersma et al., 2018). Putz et al. (2018) showed similar results for pigs kept in a “natural challenge environment”: the variance of deviations in daily feed intake and deviations in daily duration at feeder during finishing phase were positively genetically correlated to mortality and number of treatments (Putz et al., 2018), indicating that pigs with lower variance have lower mortality and need less treatments. This means that both feed intake and feed duration are indicative for health status on a genetic level, and thus shows that variation in frequently measured traits can be indicative for resilience. Also, many other studies have investigated the heritability of uniformity in production traits, which is the same as variance in deviations: reported heritabilities ranged between 0.00 and 0.15 in almost all livestock species (see Hill and Mulder, 2010 for an overview up to 2010; Neves et al., 2011; Janhunen et al., 2012; Sae-Lim et al., 2015, 2017; Mulder et al., 2016; see Elgersma et al., 2018 for more). Furthermore, the genetic coefficients of variation were generally moderate to high, which suggests a relatively large potential for genetic improvement in uniformity. These studies illustrate good potential for the use of production deviations as indicator traits for resilience.

Conditions for Measuring Resilience Indicators

For successful use of resilience indicators in breeding programs, certain conditions apply: collection of observations should be on many animals, observations need to be collected frequently

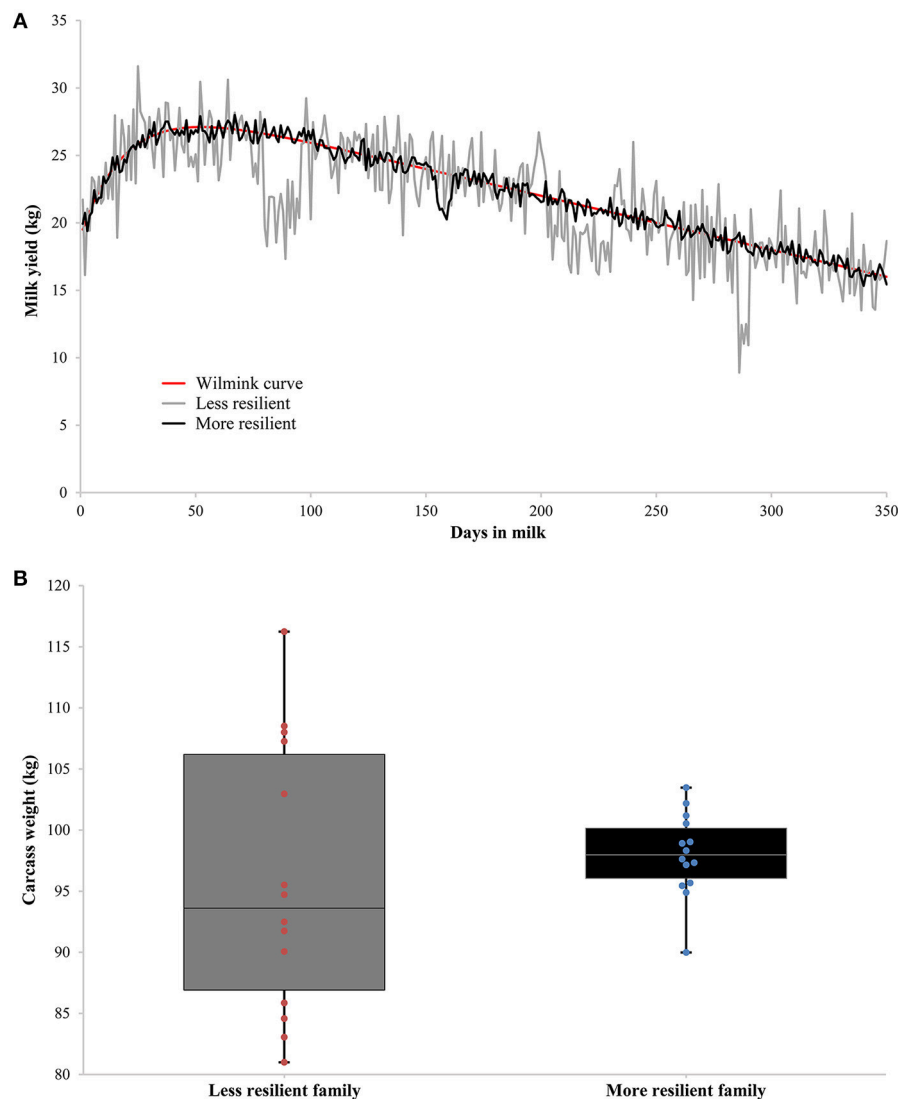


FIGURE 1 | Two examples of production traits suitable for investigating resilience indicators based on deviations. **(A)** Shows repeated measurements of milk yield on individual level of two dairy cows with the same underlying Wilmink curve (in red), but differing in resilience: in gray a less resilient dairy cow with higher fluctuations in milk yield, and in black a more resilient dairy cow with lower fluctuations in milk yield. **(B)** Shows measurements of carcass weight on family level of two families differing in resilience: in gray a less resilient family with higher fluctuations in carcass weight, and in black a more resilient family with lower fluctuations in carcass weight.

and over a longer period of time, the environment has to be challenging and diverse to estimate general resilience, and most importantly the resilience indicators have to be informative for resilience.

Production traits suitable for investigating resilience indicators based on deviations can be either measured repeatedly on individual level or on family level (see **Figure 1** for example). Repeated observations on individual level allow more accurate estimation of resilience indicators and their genetic variance and are therefore preferred. Suitable traits are for example milk yield, egg weight, body weight/growth, and feed intake. An important point that needs to be addressed is that these traits might differ in size of deviations depending on the (lactation) stage or age (for

growth of young animals): standardizing deviations per stage or age might be required or treating different stages/ages as different traits. In case too few repeated observations on production traits are collected per individual, deviations of production traits collected on (sire-)family level are an alternative. Especially livestock kept in high numbers with a relatively low economic value (e.g., fish and poultry) or livestock kept in extensive farming systems (e.g., beef cattle and sheep) can benefit from such an approach to investigate resilience indicators. Suitable traits are for example slaughter/carcass traits (e.g., Ibáñez-Escriche et al., 2008) or traits recorded only during a restricted lifetime period (e.g., Neves et al., 2011; Mulder et al., 2016; Iung et al., 2017). It is to be expected that technical developments in

the (near) future will allow investigation of deviations based on repeated observations on the same individual for all livestock species.

Care has to be taken to estimate deviations properly: less resilient animals will have more fluctuations in production, but as a consequence the expected production (based on own observations) will be lower. This means that deviations will be lower if simply an average expected production is used, and resilience of the animal is overestimated. For example, Elgersma et al. (2018) concluded that the raw phenotypic variance of daily milk yield is a composite measure of the residual variance and the shape of the lactation curve (Elgersma et al., 2018). They suggested to model the individual's lactation curve and use variance of deviations from the lactation curve as a better resilience indicator (Elgersma et al., 2018), as was done by Codrea et al. (2011). However, an individual permanently affected by disturbances (or a disturbance) will have a lactation curve lower than its true potential without (any) disturbances. As a consequence, deviations based on the individual's lactation curve will be lower than when compared to the individual's potential lactation curve. Modeling lactation curves based on the observed data might absorb part of the variance in deviations. Instead, deviations might be based on the difference between observed production and an individual's potential production, e.g., based on its (G)EBV. Determining the expected production of an animal is one of the major challenges for unbiased estimation of the proposed resilience indicators, and requires further investigation in the near future.

To properly estimate general resilience of individuals, two factors are essential: the length of the observation period and the frequency of the observations. First, the length of the observation period should be sufficiently long to allow different types of disturbances to occur, as was suggested by Mulder et al. (2013): in stressful environments more genetic variation for micro-environmental sensitivity can be observed compared to less stressful environments (Mulder et al., 2013). A potential risk is that results are not reproducible between or even within an environment: for instance, resilience based on disturbances caused by one particular type (e.g., diseases) does not reflect resilience toward other types of disturbance (e.g., heat), and therefore does not represent general resilience. Second, the frequency of observations should be sufficiently high to capture deviations caused by disturbances. Elgersma et al. (2018) suggested that earlier studies might have underestimated the potential of resilience indicators due to too long test-day intervals, e.g., monthly milk yield records. Also in case family deviations are used, there is a substantial risk that this does not capture any disturbance (in case no or small deviations were present), let alone a diverse set of disturbance types to estimate general resilience. In contrast, too small time periods may invoke larger noise, and it may therefore become harder to find the relevant information for resilience. The time period depends on the trait measured. Future research should focus on finding the optimal time period for determining deviations by looking at the accuracy of EBV, and the genetic correlations between the resilience indicator(s) and the existing health traits. Regardless of that, resilience is ideally estimated in an environment that has

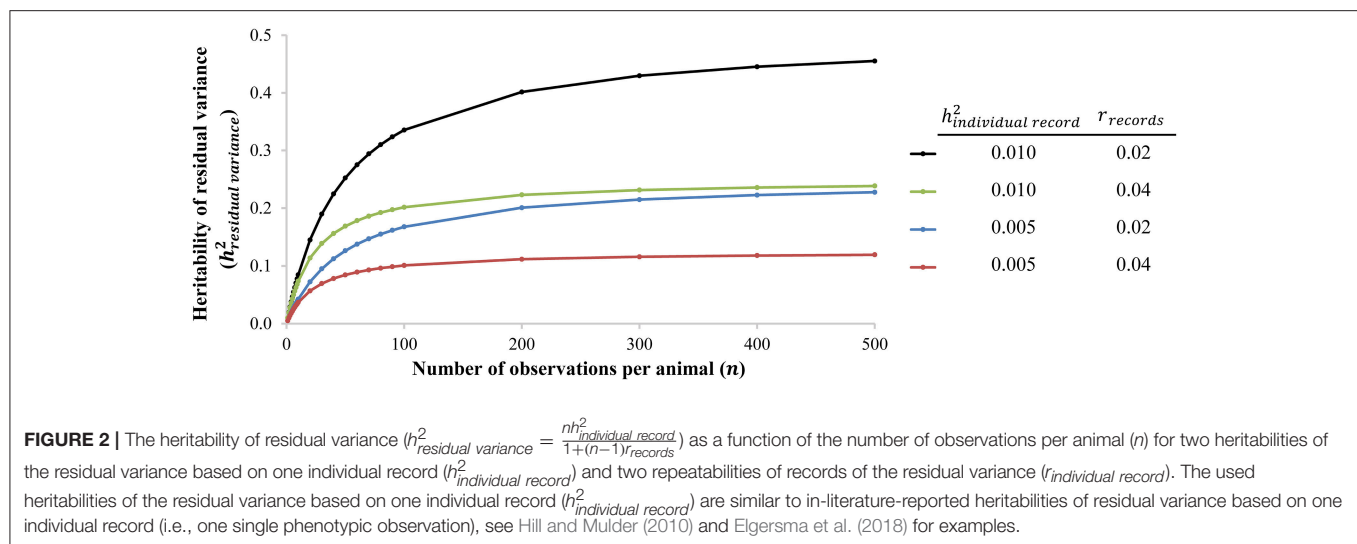
all types of disturbances for the whole production period with frequent measurements.

Recent technological developments allow a tremendous increase in number of observations and number of observed phenotypes: big/longitudinal data and new phenotypes will result in more and new data on individuals to more accurately estimate deviations and consequently genetic parameters. Currently, many (breeding) organizations make use of routine data collection. Automatic milking systems (AMS) and automatic feeding systems (AFS) for cattle and pigs are the most well-known and well-developed examples. Other developments currently under investigation are for example automatic weighing systems and automatic recording of egg production in group housing. Collection of more data and new phenotypes are expected to increase in the near future, which will aid in (better) estimation of genetic variation of deviations: higher heritabilities are generally found with more observations (i.e., individual measurements) per individual or family (Hill and Mulder, 2010; Elgersma et al., 2018), suggesting that large datasets are required to accurately estimate genetic variance. In addition, even though heritability estimates of variance of deviations based on single records are often very low (around 0.01; see Hill and Mulder, 2010; Elgersma et al., 2018 for examples), the heritability of residual variance of many of these observations of one animal is actually moderate (Damgaard et al., 2003; Kapell et al., 2011; Sell-Kubiak et al., 2015; Mulder et al., 2016; Elgersma et al., 2018). **Figure 2** shows the relationship between the heritability of the residual variance based on multiple individual records and the heritability of the residual variance based on one individual record. In summary, recent technological developments, such as AMS and AFS, allow or will allow collection of large datasets with more observations per animal (i.e., big data). This will greatly facilitate the use of resilience indicators in breeding programs of all livestock species.

THE IMPORTANCE OF RESILIENCE IN THE BREEDING GOAL

Determining the breeding goal is one of the most important elements of animal breeding. The breeding goal and the corresponding selection index determine the direction in which genetic improvement should take place. In all species, breeding goals have moved from primarily production-driven breeding goals to balanced breeding goals that aim for simultaneous improvement of production, efficiency, and health and functional traits (Olesen et al., 2003; Knap, 2009a; Neeteson-Van Nieuwenhoven et al., 2013; Miglior et al., 2017). Genetic improvement of resilience fits within the philosophy of balanced breeding, but, as far as we know, resilience is not (yet) included in breeding goals of livestock. If resilience has an impact on farm profit it should be in the breeding goal. In other words, if the economic value of resilience is nonzero, it should be a breeding goal trait. The question is, however, how can we determine the economic value for resilience.

To determine the economic value of resilience, we can consider the costs of a lack of resilience, for example higher production losses, (labor) costs of health treatments, veterinary



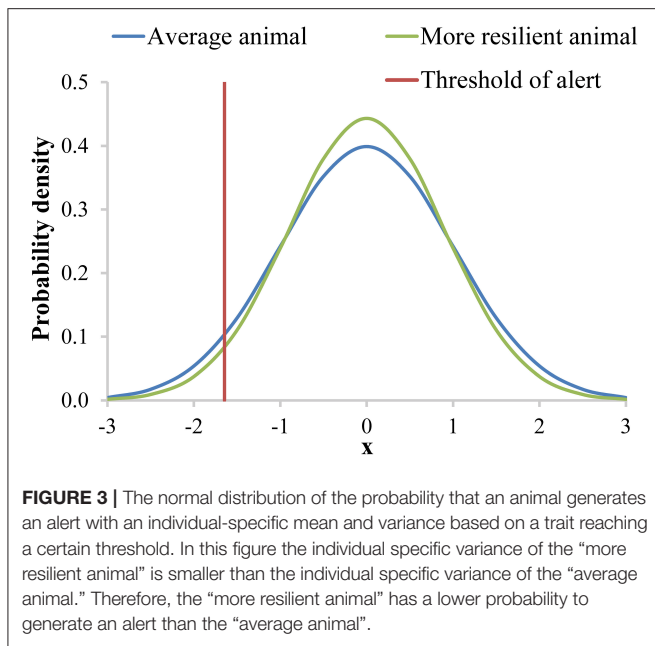
costs, and labor costs of the farmer for observing animals that show signs of lack of resilience. When determining the economic value of traits, it is important to avoid double counting. If resilience is defined as fluctuations in production, care needs to be taken to avoid double counting. For example, production losses (i.e., deviations due to a lack of resilience) might already be captured by the production traits, especially in dairy cows. Furthermore, costs of health treatments might already be accounted for in the breeding goal, for instance in the case of mastitis in dairy cattle. Treatment costs of mastitis, or costs of discarded milk should not be accounted for in resilience. On the other hand, production losses and costs of diseases are not always included in the breeding goal of, for instance, pigs and poultry: there is a lack of health traits in these breeding goals, and observed production losses in commercial or crossbred environments due to disturbances may not be observed in the high health selection environment. Therefore, the economic value of resilience in pigs and poultry may include production losses in the commercial environment and health costs. Anyway, the economic value can, for example, be based on labor costs for observing animals that show signs of disease or other problems, e.g., alerts or visual signs. These costs are often overlooked, because it is considered to be part of day-to-day management. However, if the number of animals per farm employee is increasing and labor time is restricted, this is clearly associated with the farmer's requirement for healthy and easy-to-manage animals. Genetic improvement in resilience would reduce labor requirements and would allow the farmer to keep more animals. Resilience should therefore be included in breeding goals.

For sustainable animal breeding, environmental and societal concerns have to be taken into account, in addition to economic concerns (Olesen et al., 2000; Nielsen and Amer, 2007). In other words, in addition to an economic value for resilience, a non-economic value might be present. Non-economic values can require extensive work to determine (e.g., Nielsen et al., 2011; Grimsrud et al., 2013), and could for example be based on

improved health and welfare of animals, and job satisfaction of farmers. However, for the sake of merely illustrating resilience as a concept, we will not take non-economic values into account in this paper.

Next, we will use the example of labor costs to show the potential of including resilience in breeding goals. Additional labor costs for (lower) resilience are related to the probability that the animal generates an alert. An alert is a warning that might indicate that an animal is influenced by a disturbance. An alert can be generated either by visual inspection, or by sensors, AMS or AFS. We assume that alerts are generated when a trait (with a normal distribution and an individual-specific variance) exceeds a fixed threshold value that is based on the population variance (e.g., a threshold that belongs to the population-wide 1% probability). This trait could be for instance milk yield or body weight, but likely not the previously proposed resilience indicators. The proposed resilience indicators are expected to detect problems too late, and we propose to use them to breed for more resilient animals rather than to use them as predictors of alerts. Breeding for more resilient animals will result in offspring with a smaller variance in their sensor values than the current generation, and thus a smaller probability of generating an alert (Figure 3).

We will illustrate the example of using labor costs to derive economic values in two situations: in case labor time is unrestricted, and, more realistically, in case labor time is restricted per farm (i.e., maximal number of animals given farm conditions). In addition, we will show how a combined economic value for resilience can be determined based on multiple components, in this case simplified to labor costs and (health) treatment costs. For simplicity, we choose variance as the resilience indicator for (further) illustration. We also assume that breeding for a lower variance will decrease the probability of generating alerts, and that the genetic correlation between the variance and resilience in the breeding goal is 1. We hypothesize that breeding for more resilient animals will result in offspring with an autocorrelation closer to zero, a skewness closer to zero,



and a slope closer to zero. In all cases, the offspring are expected to have a lower probability of generating an alert, because the underlying concept is the same, although it is less obvious than with a smaller variance.

Economic Value for Resilience Based on one Component

If labor time is unrestricted, the economic value of resilience is the change in expected number of alerts multiplied by the time per alert and the labor cost per time unit over the whole production cycle of an individual; i.e., using Equation 8b from Mulder et al. (2008), the economic value of resilience ($v_{\text{resilience,unrestricted labor time}}$) is:

$$v_{\text{resilience,unrestricted labor time}} = 0.5 \times z \times x \times l_a \times c_l \times d \quad (1)$$

where z is the ordinate of the standard normal distribution at the standardized threshold x of the alert (e.g., the threshold that belongs to the 1% probability), l_a is the labor time required for dealing with the alert, c_l is the labor cost per time unit, and d is the number of days of the finisher period.

If labor time is restricted per farm (i.e., the number of animals on a farm is maximized given a certain farm management), we assume that the total available time (L) is constant, and average time available per animal for normal management (l_n) cannot be changed. We also assume that the average time per animal required for dealing with alerts ($l_r = l_a \times p_a$, with p_a being the probability of obtaining an alert) can be changed by selection, i.e., selection for resilience reduces the number of alerts per animal over the whole production cycle. The total profit of a farm is:

$$\text{Profit} = n \times (\text{Revenues} - \text{Costs}) \quad (2)$$

where n is the number of animals per farm, equal to:

$$n = \frac{L}{l_n + l_r} \quad (3)$$

Rewriting Equation (2) using Equation (3) results in:

$$\text{Profit} = \frac{L}{l_n + l_r} \times (\text{Revenues} - \text{Costs}) \quad (4)$$

To obtain the economic value, the derivative of Equation 4 with respect to l_r is required, being:

$$\frac{d\text{Profit}}{dl_r} = \frac{-L}{(l_n + l_r)^2} \times (\text{Revenues} - \text{Costs}) \quad (5)$$

The economic value must be expressed per animal and therefore Equation 5 is divided by Equation 3 to obtain the improvement in profit when l_r changes with 1 time unit:

$$\frac{d\text{Profit}}{dl_r/\text{animal}} = \frac{-(\text{Revenues} - \text{Costs})}{l_n + l_r} \quad (6)$$

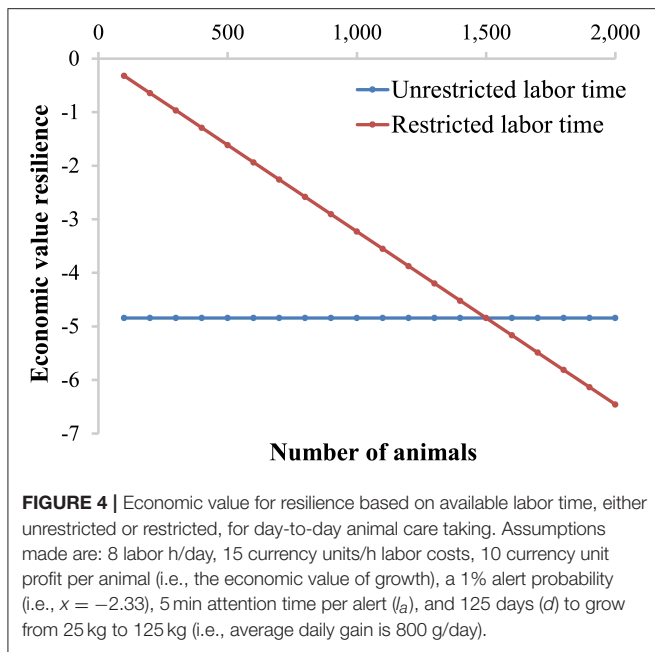
In other words, Equation (6) is the change in profit when changing l_r with 1 time unit, and can be interpreted as the cost-price of labor spent on dealing with alerts for the total period an animal is kept, i.e., the product $c_l \times d$ in Equation (1). In fact, the economic value shows the increase or decrease in farm profit due to higher or lower resilience, because more or fewer animals can be kept on the farm, if labor is restricted.

To obtain the economic value on the basis of a difference in resilience, Equation 1 (unrestricted labor time) is adjusted based on the different cost-price of restricted labor time of Equation 6. This results in:

$$v_{\text{resilience,restricted labor time}} = 0.5 \times z \times x \times l_a \times \frac{-(\text{Revenues} - \text{Costs})}{l_n + l_r} \quad (7)$$

To show the impact of unrestricted labor time and restricted labor time per farm, we calculate the economic values for resilience for a farm with finisher pigs. We assume 8 labor h/day, 15 currency units/h labor costs, 10 currency unit profit per animal (i.e., the economic value of growth), a 1% alert probability (i.e., $x = -2.33$), 5 min attention time per alert (l_a), and 125 days (d) to grow from 25 kg to 125 kg (i.e., average daily gain is 800 g/day).

Figure 4 shows that the economic value of resilience is constant when labor time is unrestricted. However, the economic value of resilience increases with increasing farm size when labor time is restricted per farm (**Figure 4**). The two situations lead to an equal economic value when the farm size is 1,500 finisher pigs, because in that case the income of the farmer would be 15 currency units/h, equal to the price of unrestricted labor. The economic value of resilience based on restricted labor time reaches more than 60% of the economic value of growth (in our example) with a farm size of 2,000 pigs, and would keep increasing with increasing farm size. Improving resilience of animals would thus allow more animals per farm (i.e., intensification). In fact



with restricted labor time, the time for normal management (l_n) per animal decreases with an increase in number of animals and a constant amount of labor time available. Therefore, the proportion of time for alerts (l_r) increases and the proportion of time for normal time (l_n) decreases with farm size. As a consequence, an increase or decrease in resilience has a larger impact on profit of the farm when the farm size increases. Nevertheless, both situations show that the economic value of resilience would be negative, meaning that reducing deviations will have a beneficial effect (i.e., an increase) on farm profit and thus resilience should be included in the breeding goal.

Economic Value for Resilience Based on Multiple Components

In the situation above, we considered only labor costs. Next, we extend the economic value of resilience to include for instance treatment costs for a disease, in case these costs are not yet included in the breeding goal. The economic value for resilience is:

$$v_{\text{resilience}} = v_{\text{labor}} + v_{\text{treatment}} \quad (8)$$

with v_{labor} being the economic value for labor costs (based either on unrestricted or restricted labor time) and $v_{\text{treatment}}$ being the economic value for treatment costs, which can be defined as:

$$v_{\text{treatment}} = 0.5 \times z \times x \times p_{\text{treatment}} \times \text{Cost}_{\text{treatment}} \quad (9)$$

where z is the ordinate of the standard normal distribution at the standardized threshold x of the alert, (e.g., the threshold that belongs to the 1% probability), $p_{\text{treatment}}$ is the probability of a treatment given that the animal got an alert, and $\text{Cost}_{\text{treatment}}$ is the cost of the treatments. Note that in this case, treatment costs must not be part of other components of the selection index

to avoid double counting of these costs. This small extended example shows that different components, based on costs, can be relatively easily included in the economic value of resilience. As shown in **Figure 4**, the economic value of resilience can easily reach 60% of the economic value of growth, but these extra components can make the economic value even larger than the economic value of growth.

THE ADDED VALUE OF RESILIENCE IN BREEDING PROGRAMS

We will now show the added value of estimated breeding values for resilience to breed healthy and easy-to-manage animals in two livestock species: (1) a pig scenario, and (2) a dairy cattle scenario. In both cases, the selection indices will be simplified in order to draw general conclusions. Selection is based on truncation selection on the index (I) to maximize response in the breeding goal (H). A simplified genomics scheme is simulated. Calculation of the responses to selection were done in SelAction v2.1 (Rutten et al., 2002), using the principle of Dekkers (2007) to include genomic information.

Pig Scenario

In this example, we describe a simplified pig scenario, in which individuals are only selected on growth rate and the breeding goal is extended with resilience. We explore the effect of adding a resilience indicator (e.g., variance of deviations) to the selection index and assume a genetic correlation of 1 between the resilience indicator and resilience in the breeding goal. Assumptions made are:

- Resilience has a heritability of 0.15 (slightly more conservative than Putz et al., 2018). The genetic correlation of resilience and growth rate is unknown and will be set to 0.25 or -0.25 . A positive genetic correlation indicates that a lower resilience indicator is correlated to a lower growth rate and *vice versa* for a negative correlation. In other words, a positive correlation is unfavorable and a negative correlation is favorable for simultaneous improvement of both traits.
- Selection is done at sexual maturation. Selection is based on all available information at that moment: own performance, BLUP, full and half sibs, and genomic breeding values.
- Genomic breeding values have an accuracy of ~ 0.77 , based on a reference population of 10,000 animals and the number of independent chromosomal segments being 1,000 (Daetwyler et al., 2008).
- The economic value of growth rate is set to 1 and -1 and the economic value of resilience will be varied between $-1,000$ and $1,000$ to obtain an ellipse of selection responses. Note that the interesting part of the ellipse is where the response in growth rate is positive and the response in the variance is negative (i.e., improved resilience).

More information about input can be found in **Supplementary Material**.

Figure 5 shows the selection response of growth rate and resilience. The desired direction of a breeding program aimed

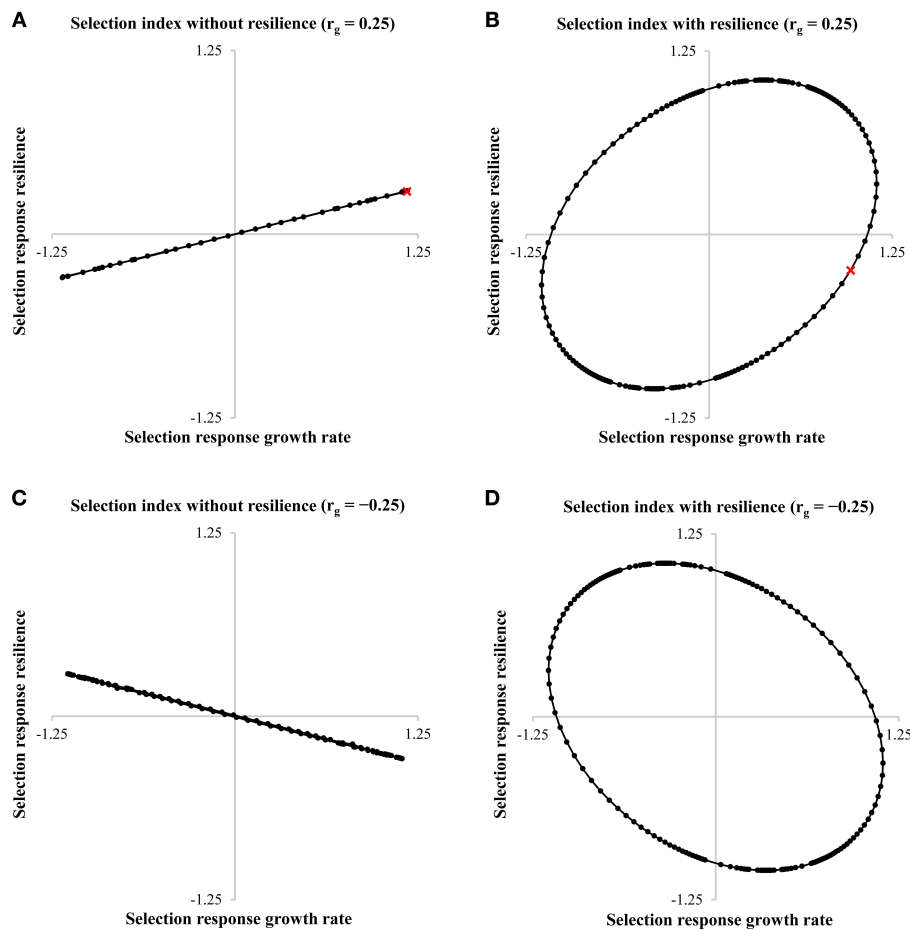


FIGURE 5 | Selection responses for growth rate and resilience with the economic value of growth rate set to 1 and -1 , and the economic value of resilience varying between $-1,000$ and $1,000$ in a pig breeding program. The figure shows the response ellipses of growth rate and resilience for the selection index without resilience (A,C) or with resilience (B,D) with a genetic correlation of 0.25 (A,B) or a genetic correlation of -0.25 (C,D). The red crosses in A,B are discussed in the text.

at simultaneously improving growth rate and resilience, i.e., reducing the variance, is the bottom right corner of each graph. Thus, in case of an unfavorable positive genetic correlation between growth rate and resilience, and resilience indicators are not included in the selection index, no progress can be made to obtain more resilient animals (Figure 5A). However, including a resilience indicator into the selection index of pigs can result in a higher selection response in the breeding goal (H) and more resilient animals, depending on the chosen economic values (Figure 5B). For example with an economic value of -0.6 for resilience (and an economic value of 1 for growth rate), the selection response in H is improved with 14.6% when a resilience indicator is included in the selection index. Although in this case, a reduction in the selection response of growth rate is observed (-15.4%), the selection response of resilience improves with 185.3% (see red crosses in Figures 5A,B). Not including a resilience indicator in the selection index increases the probability to generate an alert to 1.16% (start: 1%), while including a resilience indicator in the selection index reduces the probability to generate an alert to 0.88% . This corresponds

to a reduction of 24.7% in the number of alerts. In case of a favorable genetic correlation between growth rate and resilience, the increase in selection response in H is even higher when comparing a selection index without (Figure 5C) and with a resilience indicator (Figure 5D): $+42.8\%$ for an economic value of -1.6 for resilience. This simplified example shows that including resilience indicators in the selection index can have big impact on the selection response and number of alerts.

Dairy Cattle Scenario

In this example, we describe a simplified dairy cattle scenario, though with a more complex breeding program than the pig scenario: individuals are selected on milk yield and health-related traits, which are in this case longevity and udder health. We explore the effect of adding a resilience indicator (e.g., variance of deviations) to the selection index, i.e., selecting on a resilience indicator. Assumptions made are:

- Resilience has a heritability of 0.10 (Elgersma et al., 2018). The genetic correlation between resilience and milk yield is

0.61, between resilience and longevity is -0.30 , and between resilience and udder health is -0.36 . In other words, a higher resilience (i.e., a lower variance of deviations) is genetically correlated with a lower milk yield, a higher longevity, and a better udder health. These estimates were obtained from Elgersma et al. (2018) and CRV (2015).

- Selection is based on genomic breeding values only. Genomic breeding values have an accuracy of ~ 0.79 , based on a reference population of 20,000 animals and the number of independent chromosomal segments being 1,200 (Daetwyler et al., 2008).
- The economic values are set to 0.3 for milk yield, 0.3 for longevity, and 0.2 for udder health. In the default situation, the economic value of resilience is set to -0.2 . This means that 30% emphasis is placed on milk yield, 30% on longevity, 20% on udder health and 20% on resilience, when ignoring correlations between traits (e.g., Miglior et al., 2005). The economic value of resilience will be varied between -0.5 and -0.001 (i.e., aiming for reduced variation in deviations) as a sensitivity analysis.

More information about input can be found in **Supplementary Material**.

Similar to the pig scenario, including resilience into the breeding goal of dairy cattle can result in a higher selection response in H compared to not including resilience into the breeding goal, depending on the chosen economic values (**Figure 6**), even though health-related traits (i.e., longevity and udder health) are already included in the breeding goal and selection index. If resilience, being the variance in this case, has an economic value of -0.2 , the selection response in H increases with 3.0% (see black cross in **Figure 6**): including a resilience indicator in the selection index compensates the loss in milk yield (6.3%) by an improvement in longevity (1.4%), udder health (1.0%), and resilience (-102.6%) (**Table 2**). Not including a resilience indicator in the selection index reduces the probability to generate an alert to 0.92% (start: 1%), while including a resilience indicator in the selection index reduces the probability to generate an alert to 0.84%. This corresponds to a reduction of 8.4% in the number of alerts (see red cross in **Figure 6**). This simplified example shows that resilience indicators can have beneficial impact. However, the effect is smaller in the dairy cattle scenario compared to the pig scenario, because of the presence of health-related traits in the selection index.

Perspectives and Other Livestock Species

The potential of resilience in breeding goals was clearly illustrated in the two scenarios. Obviously these scenarios overestimate the impact of resilience indicators, because of their simplification. Nevertheless, the underlying idea holds, because a reduction of time spent on an animal with an alert (for any reason) will reduce costs and consequentially increase farm profit. The pig scenario was based on only two traits, but might still be fairly close to a pig sire line scenario. Sire line breeding programs are primarily focused on improvement of production traits, in contrast to dam line breeding programs, which are more focused on improvement of reproduction and maternal traits.

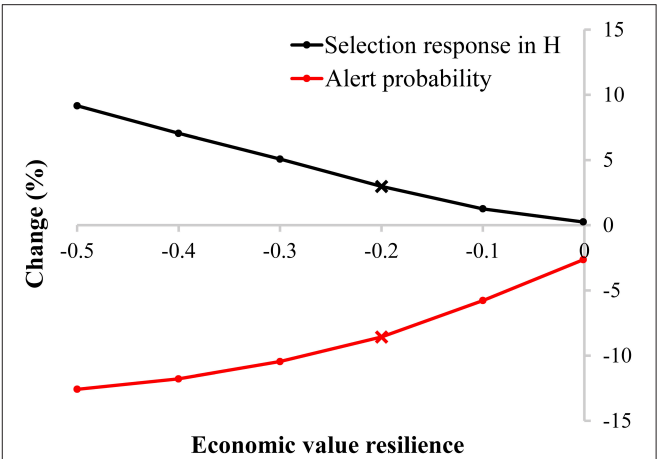


FIGURE 6 | Change (in %) in selection response in the breeding goal (H) and the probability an animal generates an alert in a default selection index without and with inclusion of a resilience indicator for various economic values in a dairy cattle breeding program. The default selection index contains milk yield, longevity, and udder health. The crosses are discussed in the text and shown in detail in **Table 2**.

TABLE 2 | Selection responses (in trait units) and alert probability (%) and their relative changes (in %) in a selection index without and with inclusion of a resilience indicator (economic value = -0.2) in a dairy cattle breeding program.

| Trait | | Selection index | | Change (%) |
|----------------------------------|-------------------------|--------------------|-----------------|------------|
| | | Without resilience | With resilience | |
| Selection response (trait units) | Milk yield | 0.80 | 0.75 | -6.3 |
| | Longevity | 1.25 | 1.27 | 1.4 |
| | Udder health | 0.80 | 0.80 | 1.0 |
| | Resilience ^a | -0.15 | -0.31 | -102.6 |
| | Breeding goal (H) | 0.80 | 0.83 | 3.0 |
| Alert probability | | 0.92% | 0.84% | -8.4 |

The default selection index contains milk yield, longevity, and udder health. Selection responses are shown for milk yield, longevity, udder health, resilience, and the breeding goal (H). ^aIn order to improve resilience of an animal, the resilience indicator trait (e.g., variation in deviations) has to be reduced. Negative selection response are therefore desired, and more negative values indicate higher resilience.

Nevertheless, also for dam lines resilience can be included in breeding programs: multiplier farms require resilient sows and resilient piglets. For instance, litter size uniformity has a beneficial effect on piglet resilience based on survival (Damgaard et al., 2003; Mulder et al., 2015b). Favorable correlations were found between the residual variance of feed intake and feed duration with mortality and the number of health treatments in pigs in a challenge environment (Putz et al., 2018). This shows that the residual variance of feed intake and feed duration can be used to improve resilience. Dairy cattle breeding programs contain health(-related) traits, which are expected to partly cover resilience indicators. Indeed, several favorable r_g between health traits and variance in deviations of milk yield were found, but none of them is equal to 1 indicating that the raw variance

of milk yield contains new information about resilience and health (Elgersma et al., 2018). This was also shown in the dairy cattle scenario, which showed considerable improvement in response to a selection index with resilience. We propose that inclusion of resilience to selection indices of livestock breeding programs (similar to pigs and dairy cattle) will strongly increase the response in resilience of livestock, and likely increase the selection response in H as well.

We did not investigate resilience in breeding programs of other livestock, such as poultry, extensively kept livestock species (e.g., beef cattle and sheep), or aquaculture species, because resilience is more difficult to assess and apply in practice at this moment. In the first place, because repeated measurements are difficult to collect. This is mainly due to the impossibility to measure animals individually, due to group housing or difficulties catching individuals. In the second place, labor time spent on alerts is less relevant. Alerts are created for some livestock species based on group measurements (e.g., water intake), but not at the individual level. Also these type of alerts represent epidemic disturbances, rather than day-to-day endemic disturbances. However, resilience indicators based on a relatively limited set of production data (both frequency of repeated observations and number of animals) can already provide valuable information on health of animals (e.g., 4-weekly body weight deviations; Berghof et al., in preparation), which is currently not incorporated into the selection indices. More importantly, the development of new techniques in the near future will allow collection of daily observations on an individual level, such as individual laying nests, individual measurements of fish (without capturing), automatic collection of data, and individual tracking and measurements with camera or drones. This might result in collection of big data and the definition of new phenotypes, and can eventually result in the use of resilience indicators in breeding programs for all livestock species.

CONCLUSION

This paper shows that including resilience in breeding programs has great potential to obtain healthy and easy-to-manage livestock. Resilience indicators can be based on deviations between observed production and expected production. Of

particular interest are variance of deviations, autocorrelation of deviations, and skewness of deviations. Also the slope of the reaction norm might contain information, though limited to macro-environmental disturbances. An economic value for resilience indicators in the selection index can be determined based on reduced labor costs and health costs, provided that these costs are not accounted for in other traits in the selection indices. For most farms, where labor time is restricted, the economic value of resilience increases with an increasing number of animals per farm. This paper also shows the additional benefit of including resilience in the breeding goal: in both the pig and dairy cattle scenarios, we show improvements in the selection response in the breeding goals and in particular the improvement of resilience by including resilience in the breeding goal. The rapid technological development on massive collection of data (i.e., big data) is only expected to increase in the near future, resulting in more data available. The accompanying possibilities to utilize these data to determine resilience indicators, will greatly facilitate breeding for improved resilience in all livestock species.

AUTHOR CONTRIBUTIONS

TB and HM conceived the project. TB, MP, and HM developed the ideas about measuring resilience. HM derived the economic value of resilience, and set up the breeding scenarios. TB performed the simulations in SelAction. TB wrote the (first draft of) the paper. MP and HM contributed to writing the paper.

FUNDING

This research was financially supported by Netherlands Organization for Scientific Research Earth and Life Sciences (NWO-ALW; project ALWSA.2016.4), and by the Dutch Ministry of Economic Affairs (TKI Agri & Food project 16022) and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs Norsvin.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00692/full#supplementary-material>

REFERENCES

- Bishop, S. C. (2012). A consideration of resistance and tolerance for ruminant nematode infections. *Front. Genet.* 3:168. doi: 10.3389/fgene.2012.00168
- Bohmanova, J., Misztal, I., and Cole, J. B. (2007). Temperature-humidity indices as indicators of milk production losses due to heat stress. *J. Dairy Sci.* 90, 1947–1956. doi: 10.3168/jds.2006-513
- Carabaño, M. J., Ramón, M., Díaz, C., Molina, A., Pérez-Guzmán, M. D., and Serradilla, J. M. (2017). BREEDING AND GENETICS SYMPOSIUM: Breeding for resilience to heat stress effects in dairy ruminants. A comprehensive review. *J. Anim. Sci.* 95, 1813–1826. doi: 10.2527/jas2016.1114
- Codrea, M. C., Hojsgaard, S., and Friggens, N. C. (2011). Differential smoothing of time-series measurements to identify disturbances in performance and quantify animal response characteristics: an example using milk yield profiles in dairy cows. *J. Anim. Sci.* 89, 3089–3098. doi: 10.2527/jas.2010-3753
- Colditz, I. G., and Hine, B. C. (2016). Resilience in farm animals: biology, management, breeding and implications for animal welfare. *Anim. Prod. Sci.* 56, 1961–1983. doi: 10.1071/AN15297
- CRV (2015). *Statistical Indicators, E-20, NVI*. Available online at: <https://www.crv4all-international.com/wp-content/uploads/2016/03/E-20-NVI.pdf>
- Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* 3:e3395. doi: 10.1371/journal.pone.0003395
- Damgaard, L. H., Rydhmer, L., Løvendahl, P., and Grandinson, K. (2003). Genetic parameters for within-litter variation in piglet birth weight and change in within-litter variation during suckling. *J. Anim. Sci.* 81, 604–610. doi: 10.2527/2003.813604x

- De Jong, G., and Bijma, P. (2002). Selection and phenotypic plasticity in evolutionary biology and animal breeding. *Livest. Sci.* 78, 195–214. doi: 10.1016/S0301-6226(02)00096-9
- Debat, V., and David, P. (2001). Mapping phenotypes: canalization, plasticity and developmental stability. *Trends Ecol. Evol.* 16, 555–561. doi: 10.1016/S0169-5347(01)00266-2
- Dekkers, J. C. M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.* 124, 331–341. doi: 10.1111/j.1439-0388.2007.00701.x
- Doeschl-Wilson, A. B., Villanueva, B., and Kyriazakis, I. (2012). The first step toward genetic selection for host tolerance to infectious pathogens: obtaining the tolerance phenotype through group estimates. *Front. Genet.* 3:265. doi: 10.3389/fgene.2012.00265
- Elgersma, G. G., De Jong, G., Van der Linde, R., and Mulder, H. A. (2018). Fluctuations in milk yield are heritable and can be used as a resilience indicator to breed healthy cows. *J. Dairy. Sci.* 101, 1240–1250. doi: 10.3168/jds.2017-13270
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to quantitative genetics (fourth edition)*. Harlow: Longman Group Ltd.
- Finlay, K. W., and Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* 14, 742–754. doi: 10.1071/AR9630742
- Flatt, T. (2005). The evolutionary genetics of canalization. *Q. Rev. Biol.* 80, 287–316. doi: 10.1086/432265
- Ge, L., Anten, N. P. R., Van Dijkhoorn, I. D. E., Feindt, P. H., Kramer, K., Leemans, R., et al. (2016). Why we need resilience thinking to meet societal challenges in bio-based production systems. *Curr. Opin. Environ. Sustain.* 23, 17–27. doi: 10.1016/j.cosust.2016.11.009
- Gijzel, S. M. W., Van de Leemput, I. A., Scheffer, M., Roppolo, M., Olde Rikkert, M. G. M., and Melis, R. J. F. (2017). Dynamical resilience indicators in time series of self-rated health correspond to frailty levels in older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 72, 991–996. doi: 10.1093/gerona/glx065
- Grimsrud, K. M., Nielsen, H. M., Navrud, S., and Olesen, I. (2013). Households' willingness-to-pay for improved fish welfare in breeding programs for farmed Atlantic salmon. *Aquaculture* 372, 19–27. doi: 10.1016/j.aquaculture.2012.10.009
- Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics* 28, 476–490.
- Herrero-Medrano, J. M., Mathur, P. K., Ten Napel, J., Rashidi, H., Alexandri, P., Knol, E. F., et al. (2015). Estimation of genetic parameters and breeding values across challenged environments to select for robust pigs. *J. Anim. Sci.* 93, 1494–1502. doi: 10.2527/jas.2014-8583
- Hill, W. G., and Mulder, H. A. (2010). Genetic analysis of environmental variation. *Genet. Res. (Camb.)* 92, 381–395. doi: 10.1017/S0016672310000546
- Holling, C. S. (1973). Resilience and stability of ecological systems. *Annu. Rev. Ecol. Syst.* 4, 1–23. doi: 10.1146/annurev.es.04.110173.000245
- Ibáñez-Escribano, N., Varona, L., Sørensen, D., and Noguera, J. L. (2008). A study of heterogeneity of environmental variance for slaughter weight in pigs. *Animal* 2, 19–26. doi: 10.1017/S1751731107001000
- Iung, L. H. S., Neves, H. H. R., Mulder, H. A., and Carvalheiro, R. (2017). Genetic control of residual variance of yearling weight in Nellore beef cattle. *J. Anim. Sci.* 95, 1425–1433. doi: 10.2527/jas.2016.1326
- Janhunen, M., Kaase, A., Vehviläinen, H., and Järvisalo, O. (2012). Genetics of microenvironmental sensitivity of body weight in rainbow trout (*Oncorhynchus mykiss*) selected for improved growth. *PLoS ONE* 7:e38766. doi: 10.1371/journal.pone.0038766
- Kapell, D. N. R. G., Ashworth, C. J., Knap, P. W., and Roehe, R. (2011). Genetic parameters for piglet survival, litter size and birth weight or its variation within litter in sire and dam lines using Bayesian analysis. *Livest. Sci.* 135, 215–224. doi: 10.1016/j.livsci.2010.07.005
- Knap, P. W. (2005). Breeding robust pigs. *Aust. J. Exp. Agric.* 45, 763–773. doi: 10.1071/EA05041
- Knap, P. W. (2009a). "Robustness," in *Resource Allocation Theory Applied to Farm Animal Production*, ed. W.M. Rauw (Wallingford: CABI Publishing), 288–301.
- Knap, P. W. (2009b). "Voluntary feed intake and pig breeding," in *Voluntary Feed Intake in Pigs*, eds D. Torrallardona and E.A. Roura (Wageningen: Wageningen Academic Publishers).
- Lade, S. J., and Gross, T. (2012). Early warning signals for critical transitions: a generalized modeling approach. *PLoS Comput. Biol.* 8:e1002360. doi: 10.1371/journal.pcbi.1002360
- Marjanovic, J., Mulder, H. A., Rönnegård, L., and Bijma, P. (2018). Modelling the co-evolution of indirect genetic effects and inherited variability. *Heredity* 121, 631–647. doi: 10.1038/s41437-018-0068-z
- Miglior, F., Fleming, A., Malchiodi, F., Brito, L. F., Martin, P., and Baes, C. F. (2017). A 100-year review: identification and genetic selection of economically important traits in dairy cattle. *J. Dairy. Sci.* 100, 10251–10271. doi: 10.3168/jds.2017-12968
- Miglior, F., Muir, B. L., and Van Doormaal, B. J. (2005). Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88, 1255–1263. doi: 10.3168/jds.S0022-0302(05)72792-2
- Mulder, H. A. (2017). Is GxE a burden or a blessing? *Opportunities for genomic selection and big data*. *J. Anim. Breed. Genet.* 134, 435–436. doi: 10.1111/jbg.12303
- Mulder, H. A., Bijma, P., and Hill, W. G. (2007). Prediction of breeding values and selection responses with genetic heterogeneity of environmental variance. *Genetics* 175, 1895–1910. doi: 10.1534/genetics.106.063743
- Mulder, H. A., Bijma, P., and Hill, W. G. (2008). Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet. Sel. Evol.* 40, 37–59. doi: 10.1051/gse:2007034
- Mulder, H. A., Herrero-Medrano, J. M., Sell-Kubiak, E., Marthur, P. K., and Knol, E. F. (2015a). "Statistical models to increase disease resilience and uniformity in animal production," in *EAAP - 66th Annual Meeting* (Warsaw)
- Mulder, H. A., Hill, W. G., and Knol, E. F. (2015b). Heritable environmental variance causes nonlinear relationships between traits: application to birth weight and stillbirth of pigs. *Genetics* 199, 1255–1269. doi: 10.1534/genetics.114.173070
- Mulder, H. A., and Rashidi, H. (2017). Selection on resilience improves disease resistance and tolerance to infections. *J. Anim. Sci.* 95, 3346–3358. doi: 10.2527/jas2017.1479
- Mulder, H. A., Rönnegård, L., Fikse, W. F., Veerkamp, R. F., and Strandberg, E. (2013). Estimation of genetic variance for macro- and micro-environmental sensitivity using double hierarchical generalized linear models. *Genet. Sel. Evol.* 45:23. doi: 10.1186/1297-9686-45-23
- Mulder, H. A., Visscher, J., and Fablet, J. (2016). Estimating the purebred-crossbred genetic correlation for uniformity of eggshell color in laying hens. *Genet. Sel. Evol.* 48:39. doi: 10.1186/s12711-016-0212-2
- Neeteson-Van Nieuwenhoven, A.-M., Knap, P., and Avendaño, S. (2013). The role of sustainable commercial pig and poultry breeding for food security. *Anim. Front.* 3, 52–57. doi: 10.2527/af.2013-0008
- Neves, H. H. R., Carvalheiro, R., Roso, V. M., and Queiroz, S. A. (2011). Genetic variability of residual variance of production traits in Nellore beef cattle. *Livest. Sci.* 142, 164–169. doi: 10.1016/j.livsci.2011.07.010
- Nielsen, H. M., and Amer, P. R. (2007). An approach to derive economic weights in breeding objectives using partial profile choice experiments. *Animal* 1, 1254–1262. doi: 10.1017/S1751731107000729
- Nielsen, H. M., Olesen, I., Navrud, S., Kolstad, K., and Amer, P. (2011). How to consider the value of farm animals in breeding goals. A review of current status and future challenges. *J. Agric. Environ. Ethics* 24, 309–330. doi: 10.1007/s10806-010-9264-4
- Olesen, I., Gjedrem, T., Bentsen, H. B., Gjerde, B., and Rye, M. (2003). Breeding programs for sustainable aquaculture. *J. Appl. Aquac.* 13, 179–204. doi: 10.1300/J028v13n03_01
- Olesen, I., Groen, A. F., and Gjerde, B. (2000). Definition of animal breeding goals for sustainable production systems. *J. Anim. Sci.* 78, 570–582. doi: 10.2527/2000.783570x
- Putz, A. M., Harding, J. C., Dyck, M. K., Canada, P., Fortain, F., Plastow, G., et al. (2018). Novel resilience phenotypes using feed intake data from a natural disease challenge model in wean-to-finish pigs. *Front. Genet.* 9:660. doi: 10.3389/fgene.2018.00660
- Rashidi, H., Mulder, H. A., Mathur, P., Van Arendonk, J. A. M., and Knol, E. F. (2014). Variation among sows in response to porcine reproductive and respiratory syndrome. *J. Anim. Sci.* 92, 95–105. doi: 10.2527/jas.2013-6889
- Ravagnolo, O., and Misztal, I. (2000). Genetic component of heat stress in dairy cattle, parameter estimation. *J. Dairy Sci.* 83, 2126–2130. doi: 10.3168/jds.S0022-0302(00)75095-8

- Rutten, M. J. M., Bijma, P., Woolliams, J. A., and Van Arendonk, J. A. M. (2002). SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. *J. Hered.* 93, 456–458. doi: 10.1093/jhered/93.6.456
- Sae-Lim, P., Kause, A., Janhunen, M., Vehviläinen, H., Koskinen, H., Gjerde, B., et al. (2015). Genetic (co)variance of rainbow trout (*Oncorhynchus mykiss*) body weight and its uniformity across production environments. *Genet. Sel. Evol.* 47:46. doi: 10.1186/s12711-015-0122-8
- Sae-Lim, P., Kause, A., Lillehammer, M., and Mulder, H. A. (2017). Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation. *Genet. Sel. Evol.* 49:33. doi: 10.1186/s12711-017-0308-3
- Scheffer, M., Bolhuis, J. E., Borsboom, D., Buchman, T. G., Gijzel, S. M. W., Goulson, D., et al. (2018). Quantifying resilience of humans and other animals. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11883–11890. doi: 10.1073/pnas.1810630115
- Scheffer, M., Carpenter, S. R., Dakos, V., and Van Nes, E. H. (2015). Generic indicators of ecological resilience: inferring the chance of a critical transition. *Annu. Rev. Ecol. Evol. Syst.* 46, 145–167. doi: 10.1146/annurev-ecolsys-112414-054242
- Sell-Kubiak, E., Bijma, P., Knol, E. F., and Mulder, H. A. (2015). Comparison of methods to study uniformity of traits: application to birth weight in pigs. *J. Anim. Sci.* 93, 900–911. doi: 10.2527/jas.2014-8313
- Van der Waaij, E. H., Bijma, P., Bishop, S. C., and Van Arendonk, J. A. M. (2000). Modeling selection for production traits under constant infection pressure. *J. Anim. Sci.* 78, 2809–2820. doi: 10.2527/2000.78112809x
- Westneat, D. F., Wright, J., and Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biol. Rev. Camb. Philos. Soc.* 90, 729–743. doi: 10.1111/brv.12131

Conflict of Interest Statement: This research was partly financed by the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics, and Topigs Norsvin. Except for the financial contribution, no other shared interests (e.g., employment, consultancy, patents, products) exist between the Breed4Food partners and the authors. This paper was neither discussed nor reviewed by any of the Breed4Food partners.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Berghof, Poppe and Mulder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Assessment of Autozygosity Derived From Runs of Homozygosity in Jinhua Pigs Disclosed by Sequencing Data

Zhong Xu¹, Hao Sun¹, Zhe Zhang¹, Qingbo Zhao¹, Babatunde Shittu Olasege¹, Qiumeng Li¹, Yang Yue¹, Peipei Ma¹, Xiangzhe Zhang¹, Qishan Wang^{1*} and Yuchun Pan^{1,2*}

OPEN ACCESS

Edited by:

Andrea B. Doeschl-Wilson,
Roslin Institute, University of
Edinburgh, United Kingdom

Reviewed by:

Maria Saura,
Instituto Nacional de Investigación y
Tecnología Agraria y Alimentaria
(INIA), Spain
Silvia Teresa Rodríguez Ramilo,
Institut National de la Recherche
Agronomique (INRA), France

*Correspondence:

Qishan Wang
wangqishan@sjtu.edu.cn
Yuchun Pan
panyuchun1963@aliyun.com

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 20 June 2018

Accepted: 12 March 2019

Published: 28 March 2019

Citation:

Xu Z, Sun H, Zhang Z, Zhao Q,
Olasege BS, Li Q, Yue Y, Ma P,
Zhang X, Wang Q and Pan Y (2019)
Assessment of Autozygosity Derived
From Runs of Homozygosity in Jinhua
Pigs Disclosed by Sequencing Data.
Front. Genet. 10:274.
doi: 10.3389/fgene.2019.00274

¹ Department of Animal Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China,
² Shanghai Key Laboratory of Veterinary Biotechnology, Shanghai, China

Jinhua pig, a well-known Chinese indigenous breed, has evolved as a pig breed with excellent meat quality, greater disease resistance, and higher prolificacy. The reduction in the number of Jinhua pigs over the past years has raised concerns about inbreeding. Runs of homozygosity (ROH) along the genome have been applied to quantify individual autozygosity to improve the understanding of inbreeding depression and identify genes associated with traits of interest. Here, we investigated the occurrence and distribution of ROH using next-generation sequencing data to characterize autozygosity in 202 Jinhua pigs, as well as to identify the genomic regions with high ROH frequencies within individuals. The average inbreeding coefficient, based on ROH longer than 1 Mb, was 0.168 ± 0.052 . In total, 18,690 ROH were identified in all individuals, among which shorter segments (1–5 Mb) predominated. Individual ROH autosome coverage ranged from 5.32 to 29.14% in the Jinhua population. On average, approximately 16.8% of the whole genome was covered by ROH segments, with the lowest coverage on SSC11 and the highest coverage on SSC17. A total of 824 SNPs (about 0.5%) and 11 ROH island regions were identified (occurring in over 45% of the samples). Genes associated with reproduction (*HOXA3*, *HOXA7*, *HOXA10*, and *HOXA11*), meat quality (*MYOD1*, *LPIN3*, and *CTNBL1*), appetite (*NUCB2*) and disease resistance traits (*MUC4*, *MUC13*, *MUC20*, *LMLN*, *ITGB5*, *HEG1*, *SLC12A8*, and *MYLK*) were identified in ROH islands. Moreover, several quantitative trait loci for ham weight and ham fat thickness were detected. Genes in ROH islands suggested, at least partially, a selection for economic traits and environmental adaptation, and should be subject of future investigation. These findings contribute to the understanding of the effects of environmental and artificial selection in shaping the distribution of functional variants in the pig genome.

Keywords: pig, runs of homozygosity, inbreeding coefficients, selection, animal breeding

INTRODUCTION

Autozygosity refers to homozygosity in which the two alleles are identical by descent (IBD). It can result from several different phenomena, such as genetic drift, consanguineous matings, population bottleneck, as well as natural and artificial selection (Curik et al., 2014). Jinhua pig, as a valuable natural resource, is a well-known indigenous breed in eastern China that has evolved as a pig breed with excellent meat quality, greater disease resistance, higher prolificacy and greater adaptability to hot and humid climate (Gao et al., 2014). Due to their superior meat quality, Jinhua pigs have been used for the production of a famous ham brand called Jinhua Ham, which is a famous ham in China (Miao et al., 2009). The number of Jinhua pigs has been decreasing in the last two decades as a result of large import of Western pig breeds to improve leanness rate of pork (China National Commission of Animal Genetic Resources, 2011). A deficient control of inbreeding may lead to a reduction of the genetic variability and therefore of the effective population size (N_e), a key parameter that influences the conservation planning and determines the rate of change in the composition of a population caused by genetic drift (Charlesworth, 2009). In addition, inbreeding may also increase the frequency of autozygosity for deleterious alleles with the consequent reduction in individual performance (Ouborg et al., 2010). For these reasons, there is a growing interest in characterizing and understanding inbreeding and autozygosity in Jinhua pigs. This would help to better preserve the genetic diversity and allow long-term viability of breeding programs of this breed.

Runs of homozygosity (ROH) are contiguous homozygous segments of the genome where the two haplotypes inherited from the parents are identical (Gibson et al., 2006). The development of high-density single nucleotide polymorphism (SNP) markers to scan the genome for ROH has been proposed as a proxy for the detection of genomic regions where a reduction in heterozygosity has occurred (Howrigan et al., 2011). Nowadays, whole genome inbreeding estimated from ROH is considered as a powerful method to distinguish between recent and ancient inbreeding (Keller et al., 2011). As the expected length of a ROH is equal to $1/2G$ Morgan, where G is the number of generations since the common ancestor (Thompson, 2013), the number of generations can be inferred from the length and frequency of ROH (Howrigan et al., 2011). The autozygosity, based on ROH, can help to improve the understanding of inbreeding depression of a trait (Keller et al., 2011) and also help to identify genes associated with traits of economic interest present in these ROH island regions (Purfield et al., 2017). In addition, given the stochastic nature of recombination, the occurrence of ROH is not randomly distributed across the genome, and ROH islands across a large number of samples may be the result of selective pressure (Zavarez et al., 2015). Recently, ROH has been used to explore signatures of selection in cattle (Peripolli et al., 2018), chicken (Marchesi et al., 2018), and sheep (Mastrangelo et al., 2017), but less commonly in pig, especially Chinese indigenous pigs such as the Jinhua pigs.

In this study, we investigated the occurrence and distribution of ROH in a sample of 202 Jinhua pigs, in order to characterize

genome-wide autozygosity levels and to detect potential ROH islands that may provide insights into past events of selection in this population. In addition, other parameters to address the levels of genetic variability, including N_e and different measures of inbreeding from pedigree and genomic information, were also investigated. For that, we used genotyping by genome reducing and sequencing (GGRS) (Chen et al., 2013), which was successfully applied to evaluate genetic diversity in Chinese indigenous pig breeds in the Taihu region (Wang et al., 2015).

MATERIALS AND METHODS

Population and Sequencing

Ear tissue samples were collected from 202 Jinhua pigs (189 females and 13 males) from conservation pig farms in Zhejiang province. Those pigs were born between 2014 and 2017, with an average depth of about four generations. A commercial kit (Lifeng Biotech, Co., Ltd., Shanghai, China) was used to extract genomic DNA, and verified the integrity and purity of DNA by agarose gel electrophoresis and the A260/280 ratio. The Genomic DNA samples were genotyped using the GGRS protocol (Chen et al., 2013). Quality control (QC) of ~ 1.4 billion raw reads were performed using NGS QC Toolkit v2.3 (Patel and Jain, 2012). In this study, we mapped the clean sequencing reads to the latest released pig reference genome (Sscrofa11.1) using BWA (Li and Durbin, 2010). SNP calling was performed using SAMTOOLS v0.1.19 and the missing genotypes were imputed using BEAGLE (Howie et al., 2009; Li et al., 2009). Additional quality controls were applied following Xiao et al. (2017). These included a minimum number of samples genotyped $> 30\%$, a calling quality > 20 (99% accuracy), and a minor allele frequency (MAF) $\geq 5\%$. SNPs mapped to sex chromosomes were excluded from the analyses.

To determine novel variants in our sequence data, we compared the identified SNPs with the dbSNP data (Build 152¹). These SNPs were annotated according to the Ensembl pig gene annotation set (Ensembl release 92²) as previously reported by Wang et al. (2015).

Genetic Diversity

Observed (H_o) and expected heterozygosity (H_e) were estimated using PLINK v1.07 (Purcell et al., 2007). The N_e was estimated using SNeP v1.1 (Barbato et al., 2015). This approach simultaneously estimated historical effective population size based on the relationship between LD, N_e , and recombination rate:

$$N_{e(t)} = (4f(c_t))^{-1} (E[r_{adj}^2 | c_t]^{-1} - \alpha) \quad (1)$$

where $N_{e(t)}$ is the effective population size t generations ago, calculated as $t = (2f(c_t))^{-1}$; c_t is the recombination rate for a specific physical distance between SNPs, which was estimated using Sved and Feldman (1973); f is the Haldane mapping function built between recombination rate and genetic distance

¹<http://ftp.ncbi.nlm.nih.gov/snp>

²http://ftp.ensembl.org/pub/release-92/gtf/sus_sscrofa/

measured by Morgan; r_{adj}^2 is the LD value corrected for sample size and α is a correction for the occurrence of mutations.

Linkage disequilibrium between SNP pairs was estimated using PLINK v1.07 (Purcell et al., 2007). Haplotype blocks were obtained with a confidence intervals algorithm and with the software Haploview (Barrett, 2009), which was also used to visualize haplotype patterns.

Measure of Runs of Homozygosity

Runs of homozygosity were identified for each individual using PLINK v1.07 (Purcell et al., 2007). The default parameter `-homozyg` were used to define ROH (Peripolli et al., 2018) and the following criteria were chosen: (1) a sliding window of 50 SNPs across the genome; (2) one heterozygous and five missing calls were allowed per window to account for genotyping error; (3) the minimum number of consecutive SNPs included in a ROH was set to 100; (4) to exclude short ROH that was derived from strong LD, the minimum length for a ROH was set to 1 Mb (Purfield et al., 2012); (5) the required minimum SNP density to define a ROH was 1 SNP per 50 kb. Considering an approximate genetic distance of 1 cM each 1 Mb (Zanella et al., 2016), a minimum ROH length of 1 Mb was expected to capture inbreeding up to 50 ancestral generations.

Pedigree and Genomic Inbreeding Coefficients

Different types of inbreeding coefficients were estimated based on pedigree and genomic information. Pedigree-based inbreeding coefficients (F_{PED}) for all pigs were estimated using R package “pedigree.”

Genomic inbreeding for each animal was estimated from ROH (F_{ROH}), as the ratio of the total length of genome covered by ROH to the total length of the genome covered by SNPs or sequences, as proposed by McQuillan et al. (2008):

$$F_{ROH} = \frac{L_{ROH}}{L_{auto}}, \quad (2)$$

in which L_{ROH} is the total length of an individual's ROH and L_{auto} is the length of the autosomal genome covered by the SNPs, which was 2.26 Gb in our study. For each animal, four ROH estimates were calculated based on lengths from sequence data as the proportion of its genome: ROH > 10 Mb ($F_{ROH > 10 Mb}$), 5–10 Mb ($F_{ROH 5-10 Mb}$), 1–5 Mb ($F_{ROH 1-5 Mb}$), and ROH > 1 Mb ($F_{ROH_{all}}$), corresponding to 5 generations, 5 to 10 generations, 10 to 50 generations, and 50 generations, respectively.

In addition, three SNP-based estimates of inbreeding coefficients were calculated using the option `-ibc` from the GCTA software (Yang et al., 2011): the first estimator, F_{SNP1} , was based on the variance of the additive genotypes (VanRaden, 2008); the F_{SNP2} estimate was calculated based on the homozygous excess; the third estimator, F_{SNP3} , was calculated based on the correlation between uniting gametes (Wright, 1922). The formulae are as follows:

$$F_{SNP1} = \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - 2p_i)^2}{h_i} - 1, \quad (3)$$

$$F_{SNP2} = 1 - \frac{1}{n} \sum_{i=1}^n \frac{Y_i(2 - Y_i)}{h_i}, \quad (4)$$

$$F_{SNP3} = \frac{1}{n} \sum_{i=1}^n \frac{Y_i^2 - Y_i(1 + 2p_i) + 2p_i^2}{h_i} \quad (5)$$

Where Y_i is the number of the reference allele copies for the i -th SNP, p_i is the frequency of this allele in the sample and $h_i = 2p_i(1 - p_i)$, and n is the total number of SNPs. Note that these coefficients were corrected by the allele frequencies of the current population and they can take negative values (Yang et al., 2013), while F_{PED} and F_{ROH} ranged from 0 to 1. The inbreeding coefficients obtained by the eight methods were compared using Pearson's correlation.

Detection of Common Runs of Homozygosity

Genomic regions with reduced genetic diversity can be found in ROH islands, and high homozygosity around the ROH islands that might harbor targets of positive selection and are under strong selective pressure (Pemberton et al., 2012). To identify the genomic overlapping ROH regions, we calculated the proportion of the occurrences of a SNP in ROH by counting the number of times the SNP was detected in those ROH across individuals, and this was plotted against the position of the SNP along the chromosome. The genomic regions most commonly associated with ROH were identified by selecting the top 0.5% of the SNPs most commonly observed in ROH (Pemberton et al., 2012). A series of adjacent SNPs, merged to constitute ROH islands and genes within each ROH island, were further extracted using the BIOMART package (Durinck et al., 2005). To further analyze the functions of identified genes, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analyses were performed using DAVID 6.8³. Only terms with a p -value less than 0.05 were considered as significant and listed.

Detection of Selection Signatures Within Jinhua Pigs

To compare the selection signatures obtained from ROH, the integrated haplotype score (iHS) test was performed within Jinhua pigs. The iHS is a measure of the amount of extended haplotype homozygosity at a given SNP, designed to use phased genotypes to identify putative regions of recent or ongoing positive selection in genomes (Voight et al., 2006). The haplotype was phased using fastPHASE with default parameters (Scheet and Stephens, 2006). The derived haplotypes were then analyzed using the *rehh* v2.0 R package (Gautier et al., 2017) as previously reported by Bertolini et al. (2018). The iHS score was computed for each autosomal SNP, and values obtained were standardized so that they followed a standard normal distribution. To calculate the p -value at the genomic level, the scores for each SNP were transformed as $p_{iHS} = -\log_{10} [1 - 2|\Phi(iHS) - 0.5|]$, where

³<https://david.ncifcrf.gov/>

$\Phi(x)$ represents the Gaussian cumulative distribution function and p_{iHS} is the two sided p -value associated with the neutral hypothesis of no selection (Gautier et al., 2017). Corresponding to the threshold of 0.5% for ROH islands, the $|iHS|$ scores higher than 2.81 ($p < 0.005$) were considered as putative signatures of selection (Cardoso et al., 2018). In this study, significant iHS signals were reported only for ROH islands.

RESULTS AND DISCUSSION

A total of 166,661 informative SNPs satisfying the quality filters were obtained, 26,458 of which were identified as unreported in the pig SNP database of NCBI. The SNP density was about 1 SNP per 13.6 kb and they were equally distributed on each chromosome, with the exception of some isolated regions (Figure 1). According to the Ensembl pig gene annotation set (Ensembl release 92), 81,753 SNPs were mapped to gene regions, of which 8,092 were mapped as exons, and 4,481 were mapped as UTRs.

Genetic Diversity

The average observed (H_o) and expected (H_e) heterozygosities in this Jinhua pig population were 0.312 ± 0.070 and 0.429 ± 0.057 , respectively. The value of H_e was slightly higher than that reported by Chen et al. (2018) for Jinhua pig breed in Zhejiang province. The earliest legend regarding the Jinhua pigs may be traced back to approximately 1600 years ago (China National Commission of Animal Genetic Resources, 2011). Considering a generation interval of 1.5 years, the 1000 generations correspond to Jinhua populations 1500 years ago, approximately. The N_e was estimated from five to 1000 generations ago in our study.

The results show that N_e has decreased through time, at a faster rate at 1,000 to 970 generations ago (Supplementary Figure S1). This finding could be explained by the domestication bottleneck caused by human-driven artificial selection approximately 1500 years ago. The effective population size in the last five generations was about 88 and was about 3018 in the 1000th generation. This value was larger than that reported by Xiao et al. (2017) for pig breeds in the Taihu region of China (ranging from 47 to 71) using the same method, therefore, this may indicate higher genetic diversity in Jinhua pigs.

Pedigree and Genomic Inbreeding Coefficients Estimate

The average inbreeding coefficients estimated using different approaches are shown in Table 1. Incomplete pedigree failed to capture the influence of relatedness among founders from the base population, thus, the levels of inbreeding based on pedigree were expected to be lower than levels of inbreeding based on ROH and SNP-by-SNP (Table 1). The average F_{ROH} based on larger segments (0.041–0.053) was closer to F_{PED} (0.01) than the average F_{SNP} . This was to some extent expected, given that the pedigree depth (about four generations) is in agreement with larger ROH segments. In addition, these two coefficients vary in the same range (0–1), while SNP-by-SNP coefficients estimated here can take lower or larger values. These low average F_{ROH} values of inbreeding suggest that recent inbreeding was low. However, according to SNP-by-SNP based coefficients, which reflect deviations of the observed inbreeding from the expected values in the current population, recent inbreeding seems to be considerably higher (0.262) than that based on pedigree and ROH.

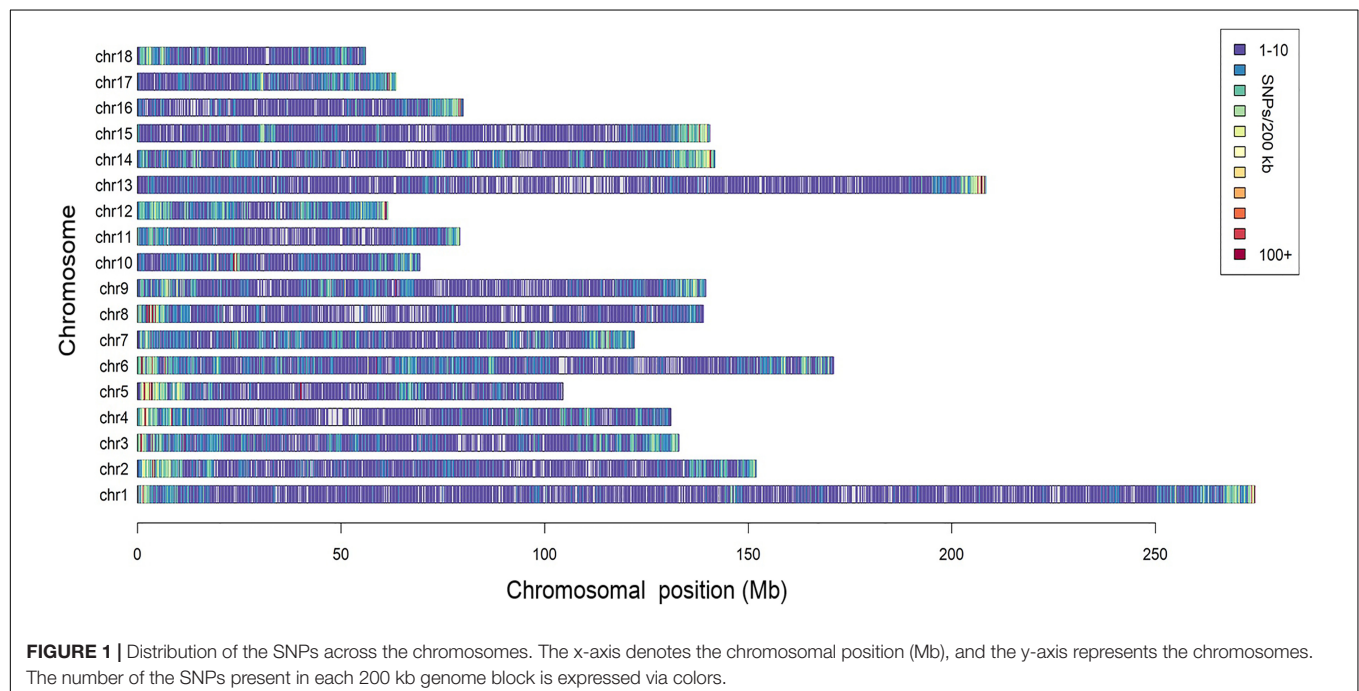


TABLE 1 | Descriptive statistics of the inbreeding coefficient based on pedigree (F_{PED}), ROH ($F_{ROH1-5\text{ Mb}}$, $F_{ROH5-10\text{ Mb}}$, $F_{ROH > 10\text{ Mb}}$, F_{ROH_all}) and SNP-by-SNP (F_{SNP1} , F_{SNP2} , F_{SNP3}).

| Inbreeding coefficient | Mean | Min | Max | SD | N |
|--------------------------|-------|--------|-------|-------|-----|
| F_{PED} | 0.010 | 0.000 | 0.137 | 0.025 | 202 |
| $F_{ROH1-5\text{ Mb}}$ | 0.073 | 0.036 | 0.118 | 0.017 | 202 |
| $F_{ROH5-10\text{ Mb}}$ | 0.041 | 0.007 | 0.076 | 0.015 | 202 |
| $F_{ROH > 10\text{ Mb}}$ | 0.053 | 0.000 | 0.157 | 0.031 | 202 |
| F_{ROH_all} | 0.168 | 0.053 | 0.291 | 0.052 | 202 |
| F_{SNP1} | 0.262 | -0.006 | 0.654 | 0.142 | 202 |
| F_{SNP2} | 0.262 | -0.148 | 0.581 | 0.181 | 202 |
| F_{SNP3} | 0.262 | -0.025 | 0.578 | 0.149 | 202 |

Min, minimum; Max, maximum; SD, standard deviation; N, the numbers of animals.

In contrast, the correlations between F_{PED} with all the genomic coefficients was low (from -0.009 to 0.053, **Table 2**), which may be indicative of a lack of power of F_{PED} to determine relatedness among founders from the base population (Visscher et al., 2006). These low correlations may also be affected by a poor and incomplete pedigree recording, as the base population assumed for F_{PED} , based on long ROH and F_{SNP} was within the range of the last four generations. Correlations between genomic coefficients, based in ROH and in SNP-by-SNP approaches were considerably higher (from 0.218 to 0.698), increasing between coefficients computed from the same source of information, as expected (i.e., ROH segments or SNP-by-SNP information), thus suggesting that genotype-based estimates provide greater accuracy on relatedness as supported by previous studies (Purfield et al., 2012; Zanella et al., 2016).

Among the four inbreeding coefficients based on different ROH lengths, F_{ROH_all} ($F_{ROH > 1\text{ Mb}}$) had higher correlations with F_{SNP1} , F_{SNP2} , and F_{SNP3} . A similar trend was also reported by Purfield et al. (2017) while studying six commercial meat sheep breeds. Among the three inbreeding coefficients based on SNP-by-SNP, F_{SNP2} had higher correlations with $F_{ROH1-5\text{ Mb}}$, $F_{ROH5-10\text{ Mb}}$, $F_{ROH > 10\text{ Mb}}$ and all F_{ROH_all} . These results corroborate previous results observed in cattle (Zhang et al., 2015a; Mastrangelo et al., 2016; Purfield et al., 2017). Similarly, Zhang et al. (2015a) also found that F_{SNP2} based

on excess of homozygosity correlated relatively highly with F_{ROH} detected from 50k and sequence data. This trend may be due to the fact that both F_{ROH} and F_{SNP2} directly reflect homozygosity on the genome (Brito et al., 2017). F_{SNP2} (to some extent) capture all of the homozygosity, whereas, the F_{ROH} uses only ROH. Furthermore, the moderate to high correlations between F_{ROH} and the three other estimates of genomic inbreeding (F_{SNP1} , F_{SNP2} , and F_{SNP3}) suggested that the proportion of the genome in ROH can be used as an accurate estimate of individual inbreeding levels (Purfield et al., 2012; Peripolli et al., 2018).

Genomic Distribution of Runs of Homozygosity

The abundance and genomic distribution of ROH provide efficient information about the demographic history of livestock species (Bosse et al., 2012). In total, 18,690 ROH were identified in 202 individuals. The mean ROH length was 4.11 Mb and the longest segment, found in chromosome SSC1 had 72.45 Mb (2,237 SNPs). The distribution of ROH according to length is shown in **Figure 2A**. The descriptive statistics of ROH number and length by classes is given in **Table 3**. The total ROH number for Jinhua pigs was composed mostly of a high number of shorter segments (1–5 Mb), which accounted for approximately 77% of all ROH detected, and contributed about 43% of the cumulative ROH length. In contrast, larger ROH (> 10 Mb), which were only 8% of all ROH, still covered about 32% of the total ROH length. These results revealed that both ancient (up to 50 generations ago) and recent (within the last five generations) inbreeding have had an impact on the genome of the Jinhua pig population.

For individuals, the relationship between total number of ROH and total length of the genome covered by ROH showed considerable variation among animals (**Supplementary Figure S2**). Individual ROH autosome coverage ranged from 5.32% (120.52 Mb) to 29.14% (659.87 Mb) in the Jinhua population. Similar distributions were also observed in other livestock species, such as sheep (Mastrangelo et al., 2017) and cattle (Peripolli et al., 2018).

For chromosomes, the number of ROH per chromosome and the percentage of chromosomes covered by ROH are shown in **Figure 2B**. The highest number of ROH per chromosome was on SSC6 (1,672 segments), whereas the lowest was on SSC16 (485 segments). On average, approximately 16.8% of the whole genome was under ROH segments, with the lowest coverage

TABLE 2 | Correlation coefficients (lower panel) between pedigree-based inbreeding coefficients (F_{PED}), four inbreeding coefficients based on different ROH lengths ($F_{ROH1-5\text{ Mb}}$, $F_{ROH5-10\text{ Mb}}$, $F_{ROH > 10\text{ Mb}}$, and F_{ROH_all}) and three inbreeding coefficients based on SNP-by-SNP (F_{SNP1} , F_{SNP2} , and F_{SNP3}).

| Correlation | F_{PED} | $F_{ROH1-5\text{ Mb}}$ | $F_{ROH5-10\text{ Mb}}$ | $F_{ROH > 10\text{ Mb}}$ | F_{ROH_all} | F_{SNP1} | F_{SNP2} | F_{SNP3} |
|--------------------------|-----------|------------------------|-------------------------|--------------------------|----------------|------------|------------|------------|
| F_{PED} | 1 | | | | | | | |
| $F_{ROH1-5\text{ Mb}}$ | 0.049 | 1 | | | | | | |
| $F_{ROH5-10\text{ Mb}}$ | 0.040 | 0.510** | 1 | | | | | |
| $F_{ROH > 10\text{ Mb}}$ | 0.042 | 0.450** | 0.453** | 1 | | | | |
| F_{ROH_all} | 0.053 | 0.752** | 0.737** | 0.885** | 1 | | | |
| F_{SNP1} | 0.029 | 0.293** | 0.218** | 0.295** | 0.339** | 1 | | |
| F_{SNP2} | -0.009 | 0.628** | 0.490** | 0.573** | 0.698** | 0.702** | 1 | |
| F_{SNP3} | 0.008 | 0.521** | 0.401** | 0.488** | 0.585** | 0.902** | 0.941** | 1 |

**Significantly different $p < 0.01$.

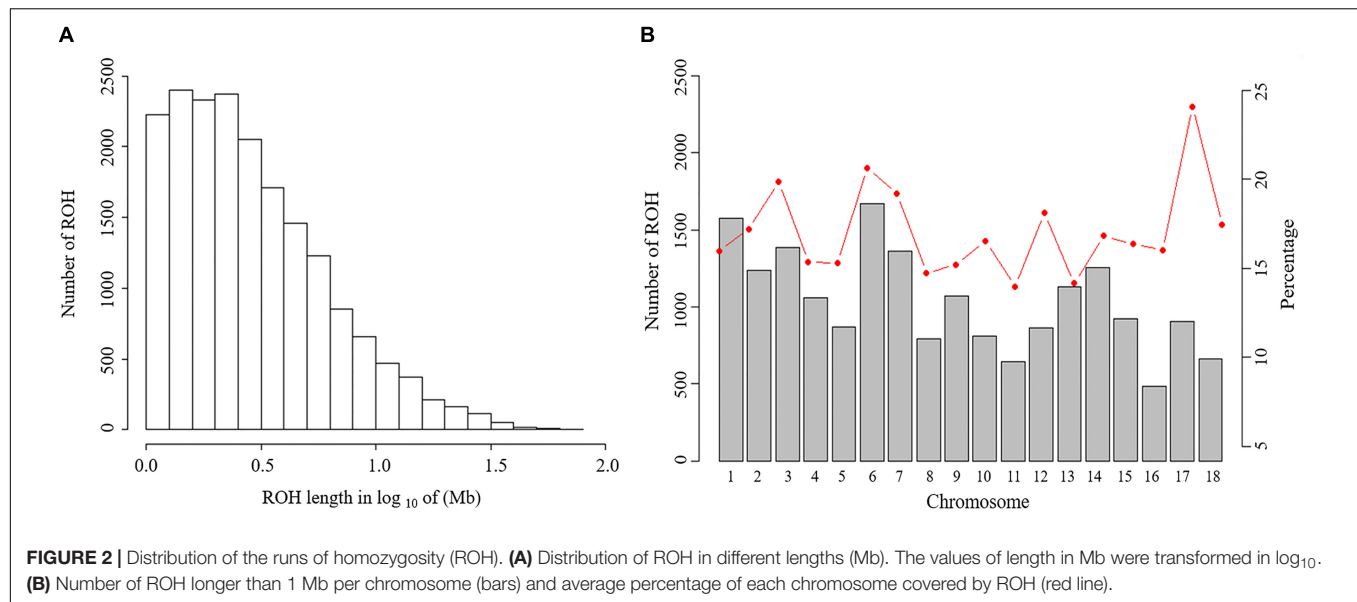


FIGURE 2 | Distribution of the runs of homozygosity (ROH). **(A)** Distribution of ROH in different lengths (Mb). The values of length in Mb were transformed in \log_{10} . **(B)** Number of ROH longer than 1 Mb per chromosome (bars) and average percentage of each chromosome covered by ROH (red line).

TABLE 3 | Descriptive statistics of runs of homozygosity (ROH) number and length (in Mb) by ROH length class (ROH 1-5 Mb, ROH 5-10 Mb, ROH > 10 Mb and total).

| ROH length (Mb) | ROH number | Percent (%) | Mean length (Mb) | Standard deviation | Genome coverage (%) |
|-----------------|------------|-------------|------------------|--------------------|---------------------|
| 1-5 | 14524 | 77.71 | 2.31 | 1.04 | 7.34 |
| 5-10 | 2753 | 14.73 | 6.85 | 1.38 | 4.12 |
| > 10 | 1413 | 7.56 | 17.24 | 7.79 | 5.33 |
| Total (> 1) | 18690 | 100.00 | 4.11 | 4.73 | 16.80 |

shown by SSC11 (14.0%) and the highest coverage of ROHs was on SSC17 (24.1%).

Detection of Runs of Homozygosity Islands

Twenty seven percent of SNPs were comprised in ROH in at least 20% of individuals, thus suggesting that candidate autozygosity regions are present in this population. This finding was similar to that reported by Ferencakovic et al. (2013). The most frequent SNP detected in ROH (131 occurrences, 64.9%) mapped at ~36 Mb in SSC3, according to the updated reference genome (Sscrofa11.1), although no genes have been currently mapped in this position, suggesting regulatory regions may be involved.

To identify the genomic regions that were most commonly associated with ROH in all individuals, the top 0.5% of SNPs with the highest occurrences (occurring in over 45% of the samples) in a ROH were considered as candidate SNPs (Figure 3). A total of 11 ROH island regions were identified, and the length of these regions ranged from 90 bp on SSC10 to 3.62 Mb on SSC3 (Table 4). The SNPs within these regions showed significantly higher linkage disequilibrium levels than the estimates obtained for the entire chromosome (Supplementary Table S1). On SSC8, we found the longest ROH cold-spot of 194 contiguous SNPs (12.56 Mb) that were not part of a ROH region in any of the individuals, thus suggesting a high heterozygosity region. This region might be produced by high recombination rates, or harboring loci with heterozygous advantage and under

selection favoring high haplotype diversity. These results are in good agreement with low LD levels determined in these regions (Supplementary Figure S3) (Barrett, 2009). In the same way, ROH islands showed high levels of LD, as expected (Supplementary Figure S4).

Candidate Genes Within Runs of Homozygosity Islands

Chromosome position, start and end position of ROH, ROH length, number of SNPs, and number of genes within the genomic regions of extended homozygosity are reported in Table 4. We found that some SNPs in ROH occurred in poor gene content regions. Some identified regions, such as that on SSC15, contained only one annotated gene, although it is longer than 1.3 Mb, either because the annotation of pig reference genome is still incomplete, or the genomic region is positioned in a non-coding region. A total number of 105 genes inside the ROH islands were analyzed using GO enrichment analysis. Supplementary Table S2 provides the chromosome position, start and end, gene name and Ensembl Gene ID for 105 genes. Supplementary Table S3 shows the significant GO terms and KEGG pathways; most of the genes were involved in metabolic pathways and biosynthetic process.

We also checked if the ROH islands overlapped with putative selection signatures in pigs in other literatures. We found that a ROH island at SSC7: 100881377-100912691 overlapped with gene *ADCK1*, involved in phosphate metabolism, which was in

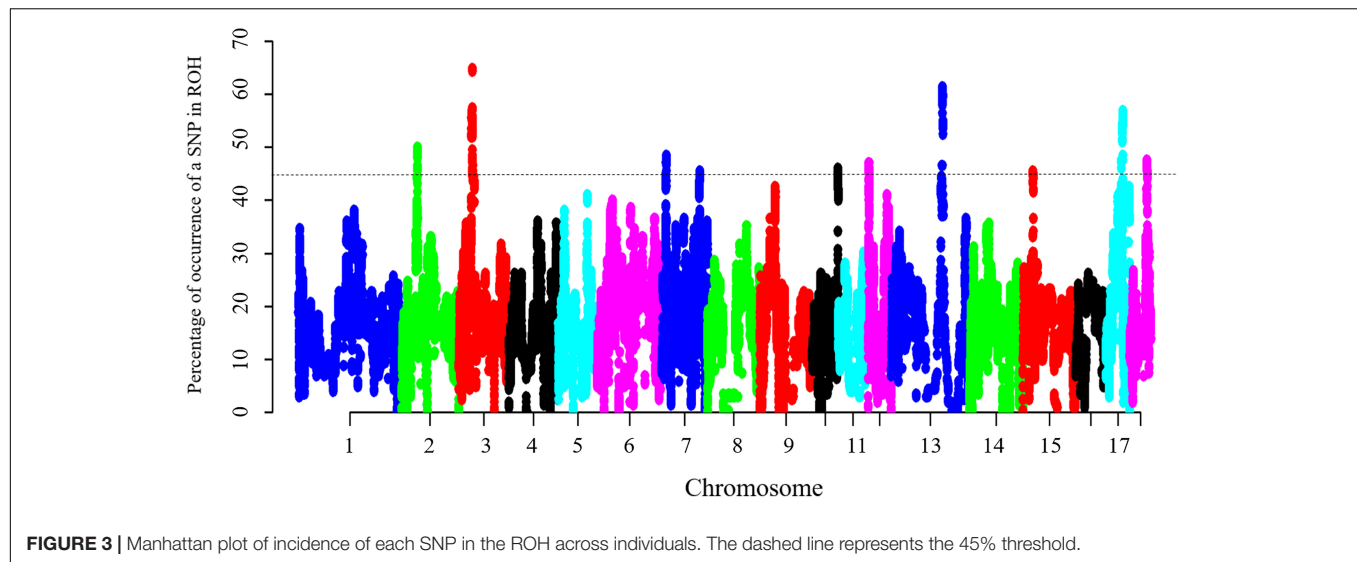


TABLE 4 | List of genomic regions of extended homozygosity detected in Jinhua pigs.

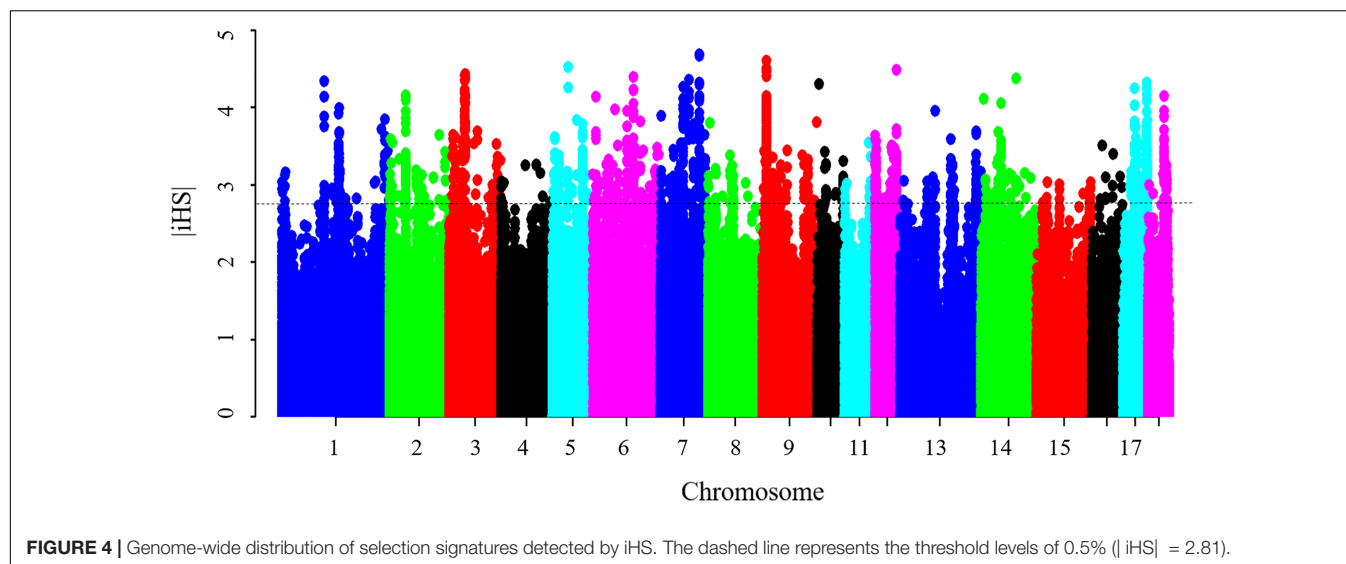
| CHR | Start (bp) | End (bp) | Length (bp) | Mean iHS value | SNPs | Genes |
|-----|------------|-----------|-------------|-------------------|------|-------|
| 2 | 41304430 | 42207694 | 903264 | 2.56 | 94 | 15 |
| 3 | 33443006 | 37063749 | 3620743 | 2.01 | 174 | 9 |
| 7 | 11825920 | 12271785 | 445865 | 1.24 | 45 | 2 |
| 7 | 100881377 | 100912691 | 31314 | 2.94 | 16 | 1 |
| 10 | 67900330 | 67900420 | 90 | 0.00 | 3 | 0 |
| 12 | 2147273 | 2619710 | 472437 | 1.51 | 59 | 12 |
| 13 | 134231214 | 137556583 | 3325369 | 1.80 | 133 | 28 |
| 15 | 25936404 | 27246878 | 1310474 | 1.24 | 62 | 1 |
| 17 | 40820642 | 40961202 | 140560 | 1.00 | 6 | 1 |
| 17 | 43470563 | 44993583 | 1523020 | 1.71 | 141 | 7 |
| 18 | 45198816 | 46247949 | 1049133 | 1.66 | 91 | 29 |

CHR, chromosome; SNPs, number of SNPs in each genomic region; Genes, number of genes in each genomic region.

the selection signature region between Berkshire and Korean native pig breeds (Edea and Kim, 2014). A ROH island on SSC2 partially overlapped with a selection region for intramuscular fat and backfat thickness in two Duroc populations, which spanned three genes (*ABCC8*, *MYOD1*, and *PIK3C2A*) (Kim et al., 2015). The *MYOD1* gene in this region was also detected in the selection signatures between Jinhua pig group and European breeds group (Li et al., 2016).

In this paper, we focused on some of the most relevant genes within ROH that showed associations with several specific traits related to livestock breeding. Several candidate genes relating to reproduction traits were identified, such as the *HOXA* genes cluster: *HOXA3*, *HOXA7*, *HOXA10*, and *HOXA11* on SSC18, which affects embryo implantation and prolificacy traits (Bagot et al., 2000; Gao et al., 2010; Wu et al., 2013); *ROPN1*, involved in litter size trait in pigs (Lan et al., 2012); and *HNRNPA2B1*, which plays key roles in the preimplantation of pig embryo during elongation (Wilson et al., 2000). Some genes associated with specific traits related to meat quality were detected: *MYOD1*, which affects muscle fiber characteristics, the loin eye area and back fat thickness

(Lee et al., 2012; Cepica et al., 2013); *LPIN3*, one member of lipin gene family associated with back-fat thickness in pigs (He et al., 2009), which are the important regulators in fat-tailed sheep with active lipid metabolism (Jiao et al., 2016); *CTNBL1*, associated with porcine fat deposition and backfat traits (Yin et al., 2012). One gene was involved in appetite: *NUCB2*, which plays an important role in whole-body energy homeostasis and body weight at puberty by regulation of appetite (Lents et al., 2013). Most of genes we detected were involved in disease resistance traits: *MUC4*, *MUC13*, *MUC20*, *LMLN*, *ITGB5*, *HEG1*, *SLC12A8*, and *MYLK* on SSC13, were potential candidate genes for controlling the expression of the enterotoxigenic *Escherichia coli* (ETEC) with F4 fimbriae (F4ac) receptor (Huang et al., 2008; Jacobsen et al., 2010; Rampoldi et al., 2011; Fu et al., 2012; Ren et al., 2012). These genes played key roles in resistance to diarrhea by defending the attachment and adhesion of ETEC to porcine jejunal cells and in maintaining the epithelial barrier as well as immunity function (Zhou et al., 2013). Many studies have revealed that resistance to ETEC F4ac adhesion in pigs can be inherited as an autosomal recessive trait, and the pigs with homozygous



genotype were usually resistant to ETEC F4ac (Rampoldi et al., 2011). The regional climate of the Jinhua pig is mainly subtropical with a weather condition that is hot and very humid, which is capable of inducing diarrhea especially during summer (Lyutskanov, 2011). Several researches have also shown that Jinhua pigs were more resistant to ETEC F4ac (Yan et al., 2009; Gao et al., 2014). These genes on SSC13 that display autozygosity in the Jinhua pigs may be linked to selection in response to hot and humid climate, as a result of local adaptation.

The Pig quantitative trait loci (QTL) database⁴ lists several QTL for reproduction, meat quality and immunity traits that overlapped these ROH islands (Quintanilla et al., 2011; Verardo et al., 2015; Zhang et al., 2016). In particular, some QTLs related to ham traits have already been reported: on SSC2, Cepica et al. (2013) identified significant QTL for ham weight (ID = 28220), Harmegnies et al. (2006) reported QTL for ham fat thickness (ID = 3938, ID = 3960, and ID = 3968); on SSC3, Choi et al. (2011) detected highly significant QTL for ham weight (ID = 21357), Stratil et al. (2006) reported QTL for ham meat weight (ID = 3102) and ham weight (ID = 3104).

In summary, the results show that the genomic regions that display autozygosity in the Jinhua pig breed are related to important production traits under selection, and possibly also help improve their adaptability to survive in hot and very humid environments.

Selection Signature Analysis

Results from the iHS test revealed that this coefficient had an average value of 0.77, with a maximum value of 4.69 on SSC7, thus indicating that the iHS values were not uniform across the genome. **Figure 4** shows the genome-wide distribution of $|iHS|$ values. The plots suggest evidence of selective forces in different regions of the genome. To compare these two

methods, the occurrence of a SNP in a ROH was correlated with the SNP $|iHS|$ value (**Supplementary Figure S5**). Significant moderate correlations were found between the iHS selection signature method and percentage of occurrence of a SNP in a ROH (Pearson's correlation coefficient = 0.25, <0.0001). The average $|iHS|$ values were also calculated in each ROH islands (**Table 4**). The result showed that mean $|iHS|$ values of SNPs in each ROH islands (except one on SSC10) were higher than that across the genome (0.77) (**Table 4**).

There were 1,535 SNPs with p -value < 0.005 that harbored signatures of selection, 92 of which were found in ROH islands (**Supplementary Table S4**). A total of 42 candidate genes were found to overlap with these regions (**Supplementary Table S5**). It includes several genes mentioned above, such as *MYOD1*, *ADCK1*, *LPIN3*, *ITGB5*, *NUCB2*, *PIK3C2A*, and *ABCC8*. These genes obtained by the two methods should be given more consideration in further studies. The significant correlation between the iHS selection signature method and the percentage of occurrence of SNP in a ROH, in the present study and elsewhere (Zhang et al., 2015b), supports the hypothesis that the observed ROH islands are not only as a result of demography, but could also be due to selection.

CONCLUSION

To our knowledge, this is the first study to describe the occurrence and distribution of ROH in the genome of Jinhua pigs. Autozygosity levels varied largely in this population, which has experienced both recent and historical inbreeding events. We have shown that, despite the low to moderate inbreeding levels in most animals, there were individuals with high inbreeding coefficients, indicating the need to account for inbreeding when planning mating strategy. Several genes within ROH islands are associated with adaptive and economic traits and should

⁴<https://www.animalgenome.org/cgi-bin/QTLdb/SS/index>

be the subject of future investigation. These findings may contribute to the understanding of the effects of environmental and artificial selection in shaping the distribution of functional variants in pig genome.

DATA AVAILABILITY

All BAM data were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). 202 samples are available under the Bioproject No. PRJNA525747.

ETHICS STATEMENT

All experimental procedures were approved by the Institutional Animal Care and Use Committee of Shanghai Jiao Tong University, and all methods involved pigs were in accordance with the agreement of Institutional Animal Care and Use Committee of Shanghai Jiao Tong University (Contract No. 2011–0033).

REFERENCES

- Bagot, C. N., Troy, P. J., and Taylor, H. S. (2000). Alteration of maternal Hoxa10 expression by in vivo gene transfection affects implantation. *Gene Ther.* 7, 1378–1384. doi: 10.1038/sj.gt.3301245
- Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6:109. doi: 10.3389/fgene.2015.00109
- Barrett, J. C. (2009). Haploview: visualization and analysis of SNP genotype data. *Cold Spring Harb. Protoc.* 2009:pdip71. doi: 10.1101/pdb.ip71
- Bertolini, F., Servin, B., Talenti, A., Rochat, E., Kim, E. S., Oget, C., et al. (2018). Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* 50:57. doi: 10.1186/s12711-018-0421-y
- Bosse, M., Megens, H. J., Madsen, O., Paudel, Y., Frantz, L. A., Schook, L. B., et al. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genet.* 8:e1003100. doi: 10.1371/journal.pgen.1003100
- Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., et al. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics* 18:229. doi: 10.1186/s12863-015-0220-1
- Cardoso, D. F., Albuquerque, L. G., Reimer, C., Qanbari, S., Erbe, M., Nascimento, A. V., et al. (2018). Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. *Genet. Sel. Evol.* 50, 22. doi: 10.1186/s12711-018-0381-2
- Cepica, S., Zambonelli, P., Weisz, F., Bigi, M., Knoll, A., Vykoukalova, Z., et al. (2013). Association mapping of quantitative trait loci for carcass and meat quality traits at the central part of chromosome 2 in Italian Large White pigs. *Meat Sci.* 95, 368–375. doi: 10.1016/j.meatsci.2013.05.002
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10:195. doi: 10.1038/nrg2526
- Chen, J., Peng, J., Xiao, Q., Pan, Y., Zhang, X., Lo, L. J., et al. (2018). The genetic diversity and population structures of indigenous pig breeds in Zhejiang Province revealed by GGRS sequencing. *Anim. Genet.* 49, 36–42. doi: 10.1111/age.12625
- Chen, Q., Ma, Y., Yang, Y., Chen, Z., Liao, R., Xie, X., et al. (2013). Genotyping by genome reducing and sequencing for outbred animals. *PLoS One* 8:e67500. doi: 10.1371/journal.pone.0067500

AUTHOR CONTRIBUTIONS

YP, QW, and ZX designed the experiments. ZX, HS, QZ, XZ, QL, and YY performed the experiments. ZX, ZZ, and PM developed some of the analysis software. ZX wrote the manuscript with the help of BO. All authors read and approved the final manuscript.

FUNDING

This work was supported by Zhejiang Province agriculture (livestock) varieties breeding Key Technology R&D Program (Grant No. 2016C02054-2), National Natural Science Foundation of China (Grant No. 31872976), and National Natural Science Foundation of China (Grant No. U1402266).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00274/full#supplementary-material>

- China National Commission of Animal Genetic Resources (2011). *Animal Genetic Resources in China Pigs*. Beijing: Chinese Agriculture Press.
- Choi, I., Steibel, J. P., Bates, R. O., Raney, N. E., Rumph, J. M., and Ernst, C. W. (2011). Identification of carcass and meat quality QTL in an F2 Duroc (Pietrain pig resource population using different least-squares analysis models. *Front. Genet.* 2:18. doi: 10.3389/fgene.2011.00018
- Curik, I., Ferencakovic, M., and Solkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., et al. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21, 3439–3440. doi: 10.1093/bioinformatics/bti525
- Edea, Z., and Kim, K. S. (2014). A whole genomic scan to detect selection signatures between Berkshire and Korean native pig breeds. *J. Anim. Sci. Technol.* 56:23. doi: 10.1186/2055-0391-56-23
- Ferencakovic, M., Solkner, J., and Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42. doi: 10.1186/1297-9686-45-42
- Fu, W. X., Liu, Y., Lu, X., Niu, X. Y., Ding, X. D., Liu, J. F., et al. (2012). A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine. *PLoS One* 7:e32127. doi: 10.1371/journal.pone.0032127
- Gao, Y., Hyttel, P., and Hall, V. J. (2010). Regulation of H3K27me3 and H3K4me3 during early porcine embryonic development. *Mol. Reprod. Dev.* 77, 540–549. doi: 10.1002/mrd.21180
- Gao, Y., Rong, Y., Wang, Y., Xiong, H., Huang, X., Han, F., et al. (2014). Expression pattern of porcine antimicrobial peptide PR-39 and its induction by enterotoxigenic *Escherichia coli* (ETEC) F4ac. *Vet. Immunol. Immunopathol.* 160, 260–265. doi: 10.1016/j.vetimm.2014.05.012
- Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* 17, 78–90. doi: 10.1111/1755-0998.12634
- Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493
- Harmegnies, N., Davin, F., De Smet, S., Buys, N., Georges, M., and Coppieiers, W. (2006). Results of a whole-genome quantitative trait locus scan for growth,

- carcass composition and meat quality in a porcine four-way cross. *Anim. Genet.* 37, 543–553. doi: 10.1111/j.1365-2052.2006.01523.x
- He, X., Xu, X., and Liu, B. (2009). Molecular characterization, chromosomal localization and association analysis with back-fat thickness of porcine LPIN2 and LPIN3. *Mol. Biol. Rep.* 36, 1819–1824. doi: 10.1007/s11033-008-9385-2
- Howie, B. N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5:e1000529. doi: 10.1371/journal.pgen.1000529
- Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12:460. doi: 10.1186/1471-2164-12-460
- Huang, X., Ren, J., Yan, X., Peng, Q., Tang, H., Zhang, B., et al. (2008). Polymorphisms of three gene-derived STS on pig chromosome 13q41 are associated with susceptibility to enterotoxigenic *Escherichia coli* F4ab/ac in pigs. *Sci. China C Life Sci.* 51, 614–619. doi: 10.1007/s11427-008-0078-9
- Jacobsen, M., Kracht, S. S., Estes, G., Cirera, S., Edfors, I., Archibald, A. L., et al. (2010). Refined candidate region specified by haplotype sharing for *Escherichia coli* F4ab/F4ac susceptibility alleles in pigs. *Anim. Genet.* 41, 21–25. doi: 10.1111/j.1365-2052.2009.01952.x
- Jiao, X. L., Jing, J. J., Qiao, L. Y., Liu, J. H., Li, L. A., Zhang, J., et al. (2016). Ontogenetic expression of Lpin2 and Lpin3 genes and their associations with traits in two breeds of chinese fat-tailed sheep. *Asian Australas. J. Anim. Sci.* 29, 333–342. doi: 10.5713/ajas.15.0467
- Keller, M. C., Visscher, P. M., and Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237–249. doi: 10.1534/genetics.111.130922
- Kim, E. S., Ros-Freixedes, R., Pena, R. N., Baas, T. J., Estany, J., and Rothschild, M. F. (2015). Identification of signatures of selection for intramuscular fat and backfat thickness in two Duroc populations. *J. Anim. Sci.* 93, 3292–3302. doi: 10.2527/jas.2015-8879
- Lan, J., Zhao, J., and Liu, Y. (2012). Molecular cloning, sequence characterization, polymorphism and association analysis of porcine ROPN1 gene. *Mol. Biol. Rep.* 39, 2739–2743. doi: 10.1007/s11033-011-1029-2
- Lee, E. A., Kim, J. M., Lim, K. S., Ryu, Y. C., Jeon, W. M., and Hong, K. C. (2012). Effects of variation in porcine MYOD1 gene on muscle fiber characteristics, lean meat production, and meat quality traits. *Meat. Sci.* 92, 36–43. doi: 10.1016/j.meatsci.2012.03.018
- Lents, C. A., Barb, C. R., Hausman, G. J., Nonneman, D., Heidorn, N. L., Cisse, R. S., et al. (2013). Effects of nesfatin-1 on food intake and LH secretion in prepubertal gilts and genomic association of the porcine NUCB2 gene with growth traits. *Domest. Anim. Endocrinol.* 45, 89–97. doi: 10.1016/j.domaniend.2013.06.002
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, Z., Chen, J., Wang, Z., Pan, Y., Wang, Q., Xu, N., et al. (2016). Detection of selection signatures of population-specific genomic regions selected during domestication process in Jinhua pigs. *Anim. Genet.* 47, 672–681. doi: 10.1111/age.12475
- Lyutskanov, M. (2011). Epidemiological characteristics of post-weaning diarrhoea associated with toxin-producing *Escherichia coli* in large intensive pig farms. *Trakia J. Sci.* 9, 68–73.
- Marchesi, J. A. P., Buzanskas, M. E., Cantao, M. E., Ibelli, A. M. G., Peixoto, J. O., Joaquin, L. B., et al. (2018). Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal* 12, 1126–1134. doi: 10.1017/S1751731117002671
- Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M. T., and Portolano, B. (2016). Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal* 10, 746–754. doi: 10.1017/S1751731115002943
- Mastrangelo, S., Tolone, M., Sardina, M. T., Sottile, G., Sutura, A. M., Di Gerlando, R., et al. (2017). Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle del Belice sheep. *Genet. Sel. Evol.* 49:84. doi: 10.1186/s12711-017-0360-z
- McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372. doi: 10.1016/j.ajhg.2008.08.007
- Miao, Z. G., Wang, L. J., Xu, Z. R., Huang, J. F., and Wang, Y. R. (2009). Developmental changes of carcass composition, meat quality and organs in the Jinhua pig and Landrace. *Animal* 3, 468–473. doi: 10.1017/S1751731108003613
- Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R., and Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends Genet.* 26, 177–187. doi: 10.1016/j.tig.2010.01.001
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. doi: 10.1371/journal.pone.0030619
- Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., and Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* 91, 275–292. doi: 10.1016/j.ajhg.2012.06.014
- Peripolli, E., Stafuzza, N. B., Munari, D. P., Lima, A. L. F., Irgang, R., Machado, M. A., et al. (2018). Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. *BMC Genomics* 19:34. doi: 10.1186/s12864-017-4365-3
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70
- Purfield, D. C., McParland, S., Wall, E., and Berry, D. P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One* 12:e0176780. doi: 10.1371/journal.pone.0176780
- Quintanilla, R., Pena, R. N., Gallardo, D., Canovas, A., Ramirez, O., Diaz, I., et al. (2011). Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects. *J. Anim. Sci.* 89, 2963–2971. doi: 10.2527/jas.2011-3974
- Rampoldi, A., Jacobsen, M. J., Bertschinger, H. U., Joller, D., Burgi, E., Vogeli, P., et al. (2011). The receptor locus for *Escherichia coli* F4ab/F4ac in the pig maps distal to the MUC4-LMLN region. *Mamm. Genome* 22, 122–129. doi: 10.1007/s00335-010-9305-3
- Ren, J., Yan, X., Ai, H., Zhang, Z., Huang, X., Ouyang, J., et al. (2012). Susceptibility towards enterotoxigenic *Escherichia coli* F4ac diarrhea is governed by the MUC13 gene in pigs. *PLoS One* 7:e44573. doi: 10.1371/journal.pone.0044573
- Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802
- Stratil, A., Van Poucke, M., Bartenschlager, H., Knoll, A., Yerle, M., Peelman, L. J., et al. (2006). Porcine OGN and ASPN: mapping, polymorphisms and use for quantitative trait loci identification for growth and carcass traits in a Meishan x Pietrain intercross. *Anim. Genet.* 37, 415–418. doi: 10.1111/j.1365-2052.2006.01480.x
- Sved, J., and Feldman, M. (1973). Correlation and probability methods for one and two loci. *Theor. Popul. Biol.* 4, 129–132. doi: 10.1016/0040-5809(73)90008-7
- Thompson, E. A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics* 194, 301–326. doi: 10.1534/genetics.112.148825
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Verardo, L. L., Silva, F. F., Varona, L., Resende, M. D., Bastiaansen, J. W., Lopes, P. S., et al. (2015). Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. *J. Appl. Genet.* 56, 123–132. doi: 10.1007/s13353-014-0240-y
- Visscher, P. M., Medland, S. E., Ferreira, M. A., Morley, K. I., Zhu, G., Cornes, B. K., et al. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2:e41. doi: 10.1371/journal.pgen.0020041
- Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72. doi: 10.1371/journal.pbio.0040072
- Wang, Z., Chen, Q., Yang, Y., Liao, R., Zhao, J., Zhang, Z., et al. (2015). Genetic diversity and population structure of six Chinese indigenous pig breeds in the Taihu Lake region revealed by sequencing data. *Anim. Genet.* 46, 697–701. doi: 10.1111/age.12349

- Wilson, M. E., Sonstegard, T. S., Smith, T. P., Fahrenkrug, S. C., and Ford, S. P. (2000). Differential gene expression during elongation in the preimplantation pig embryo. *Genesis* 26, 9–14. doi: 10.1002/(SICI)1526-968X(200001)26:1<9::AID-GENE4>3.0.CO;2-1
- Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.* 56, 330–338. doi: 10.1086/279872
- Wu, D., Song, D., Li, X., Yu, M., Li, C., and Zhao, S. (2013). Molecular characterization and identification of the E2/P4 response element in the porcine HOXA10 gene. *Mol. Cell Biochem.* 374, 213–222. doi: 10.1007/s11010-012-1522-5
- Xiao, Q., Zhang, Z., Sun, H., Wang, Q., and Pan, Y. (2017). Pudong White pig: a unique genetic resource disclosed by sequencing data. *Animal* 11, 1117–1124. doi: 10.1017/S1751731116002494
- Yan, X., Huang, X., Ren, J., Zou, Z., Yang, S., Ouyang, J., et al. (2009). Distribution of *Escherichia coli* F4 adhesion phenotypes in pigs of 15 Chinese and Western breeds and a White Duroc × Erhualian intercross. *J. Med. Microbiol.* 58, 1112–1117. doi: 10.1099/jmm.0.009803-0
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2013). “Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations,” in *Genome-Wide Association Studies and Genomic Prediction*, eds C. Gondro, J. van der Werf, and B. Hayes (Totowa, NJ: Humana Press), 215–236.
- Yin, Q., Yang, H., Han, X., Fan, B., and Liu, B. (2012). Isolation, mapping, SNP detection and association with backfat traits of the porcine CTNBL1 and DGAT2 genes. *Mol. Biol. Rep.* 39, 4485–4490. doi: 10.1007/s11033-011-1238-8
- Zanella, R., Peixoto, J. O., Cardoso, F. F., Cardoso, L. L., Biegelmeyer, P., Cantao, M. E., et al. (2016). Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genet. Sel. Evol.* 48:24. doi: 10.1186/s12711-016-0203-3
- Zavarez, L. B., Utsunomiya, Y. T., Carmo, A. S., Neves, H. H., Carnevalheiro, R., Ferencakovic, M., et al. (2015). Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes. *Front. Genet.* 6:5. doi: 10.3389/fgene.2015.00005
- Zhang, J., Chen, J. H., Liu, X. D., Wang, H. Y., Liu, X. L., Li, X. Y., et al. (2016). Genomewide association studies for hematological traits and T lymphocyte subpopulations in a Duroc × Erhualian F resource population. *J. Anim. Sci.* 94, 5028–5041. doi: 10.2527/jas.2016-0924
- Zhang, Q., Calus, M. P., Guldbandsen, B., Lund, M. S., and Sahana, G. (2015a). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16:88. doi: 10.1186/s12863-015-0227-7
- Zhang, Q., Guldbandsen, B., Bosse, M., Lund, M. S., and Sahana, G. (2015b). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16:542. doi: 10.1186/s12864-015-1715-x
- Zhou, C., Liu, Z., Liu, Y., Fu, W., Ding, X., Liu, J., et al. (2013). Gene silencing of porcine MUC13 and ITGB5: candidate genes towards *Escherichia coli* F4ac adhesion. *PLoS One* 8:e70303. doi: 10.1371/journal.pone.0070303

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Xu, Sun, Zhang, Zhao, Olasege, Li, Yue, Ma, Zhang, Wang and Pan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Genetics of Life and Death: Virus-Host Interactions Underpinning Resistance to African Swine Fever, a Viral Hemorrhagic Disease

Christopher L. Netherton^{1*}, Samuel Connell¹, Camilla T. O. Benfield² and Linda K. Dixon^{1*}

¹ The Pirbright Institute, Woking, United Kingdom, ² Royal Veterinary College, London, United Kingdom

OPEN ACCESS

Edited by:

Andrea B. Doeschl-Wilson,
Roslin Institute, University
of Edinburgh, United Kingdom

Reviewed by:

Filippo Biscarini,
Italian National Research
Council (CNR), Italy
Christine Tait-Burkard,
University of Edinburgh,
United Kingdom

*Correspondence:

Christopher L. Netherton
chris.netherton@pirbright.ac.uk
Linda K. Dixon
linda.dixon@pirbright.ac.uk

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 13 July 2018

Accepted: 12 April 2019

Published: 03 May 2019

Citation:

Netherton CL, Connell S,
Benfield CTO and Dixon LK (2019)
The Genetics of Life and Death:
Virus-Host Interactions Underpinning
Resistance to African Swine Fever,
a Viral Hemorrhagic Disease.
Front. Genet. 10:402.
doi: 10.3389/fgene.2019.00402

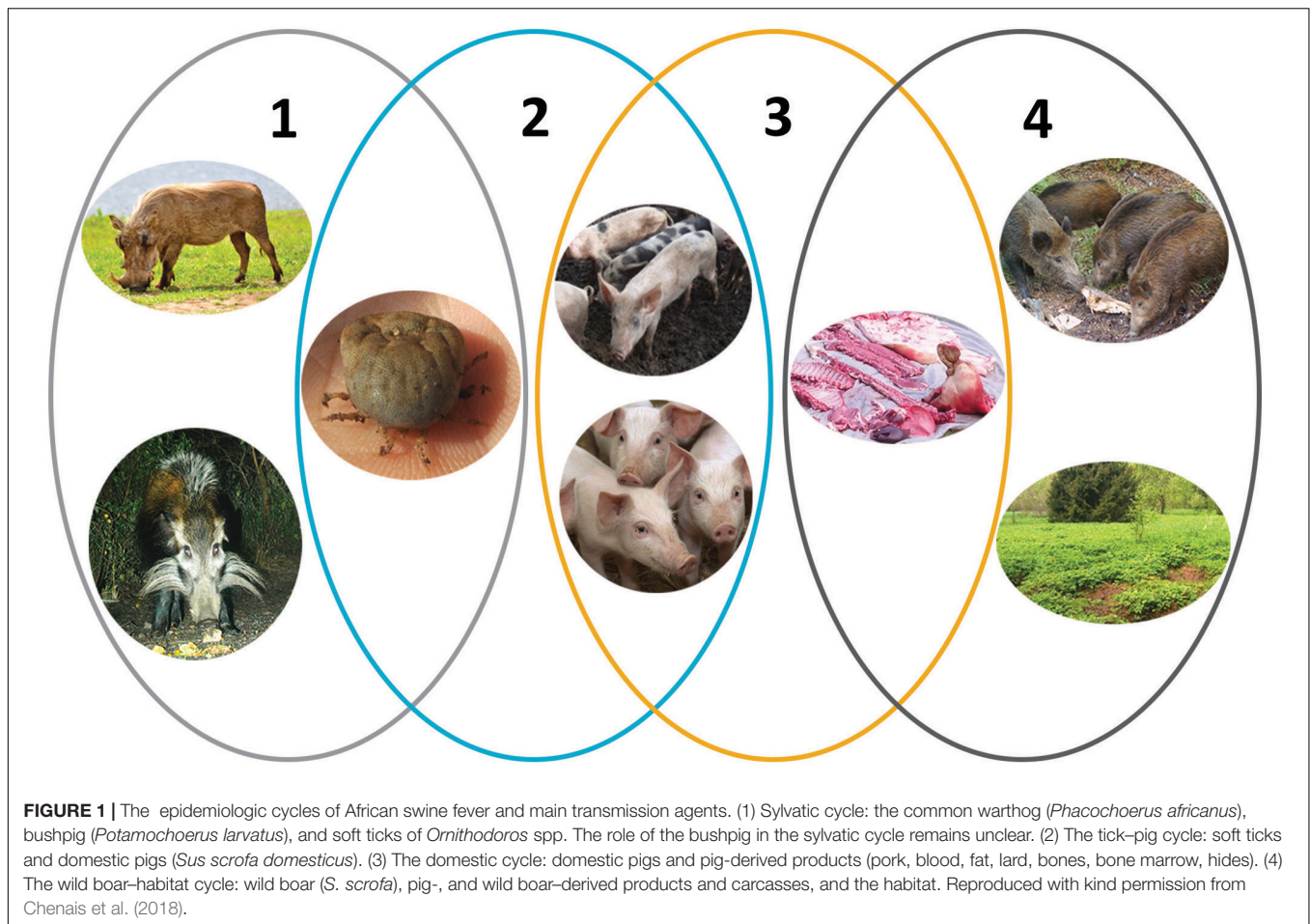
Pathogen transmission from wildlife hosts to genetically distinct species is a major driver of disease emergence. African swine fever virus (ASFV) persists in sub-Saharan Africa through a sylvatic cycle between warthogs and soft ticks that infest their burrows. The virus does not cause disease in these animals, however transmission of the virus to domestic pigs or wild boar causes a hemorrhagic fever that is invariably fatal. ASFV transmits readily between domestic pigs and causes economic hardship in areas where it is endemic. The virus is also a significant transboundary pathogen that has become established in Eastern Europe, and has recently appeared in China increasing the risk of an introduction of the disease to other pig producing centers. Although a DNA genome mitigates against rapid adaptation of the virus to new hosts, extended epidemics of African swine fever (ASF) can lead to the emergence of viruses with reduced virulence. Attenuation in the field leads to large deletions of genetic material encoding genes involved in modulating host immune responses. Therefore resistance to disease and tolerance of ASFV replication can be dependent on both virus and host factors. Here we describe the different virus-host interfaces and discuss progress toward understanding the genetic determinants of disease outcome after infection with ASFV.

Keywords: African swine fever virus (ASFV), interferon, warthog, *Ornithodoros*, host resistance, host tolerance, viral hemorrhagic fever, DNA virus infection

INTRODUCTION

African swine fever virus (ASFV) is present in a stable equilibrium with its wildlife hosts, warthogs and soft ticks of *Ornithodoros* spp., in a unique ecological niche in Eastern and Southern Africa. In these hosts virus can persist over an extended time without causing disease. However, infection of domestic pigs or wild boar with ASFV results in an invariably fatal disease, African swine fever (ASF), which is readily spread between infected pigs or wild boar without the requirement of a tick vector (**Figure 1**). ASFV can also infect and replicate in bushpigs, but like the warthog these animals

Abbreviations: ASF, African swine fever; ASFV, African swine fever virus; IFN, interferon; ISG, interferon stimulated gene; MGF, multigene family.



do not exhibit clinical signs of disease. Understanding the virus interactions with, and evolution within, these different hosts will help establish the basis for the dramatically varying pathogenesis and potentially unravel the basis for disease resistance of the wild suids in Africa.

ASF was first recognized in the early twentieth century in Kenya as an acute hemorrhagic fever that caused death of most infected domestic pigs (Montgomery, 1921). Early experiments established that warhogs did not show clinical signs of disease but provided a reservoir of infection. ASF was recognized in many Eastern and Southern African countries soon after the initial description and spread further through central and West Africa (Jori et al., 2013; Penrith et al., 2013). From Africa, ASFV expanded into Portugal in 1957 and 1960 and became endemic in the Iberian Peninsula until it was eradicated in the mid-1990s. During this time the disease also became established in Sardinia and sporadic outbreaks were reported in Western Europe, Brazil and the Caribbean. From 1999, with the exception of Sardinia, no further outbreaks of ASF were reported outside of Africa until its appearance in Georgia in the Caucasus region in 2007. Currently ASF is present in sub-Saharan Africa, Sardinia, the Trans Caucasus, the Russian Federation, and Central and Eastern states of the European Union. ASFV continues to spread, with first reports of the disease in China (August 2018),

Bulgaria (August 2018), Belgium (September 2018), and Vietnam (February 2019) highlighting the increasing threat of ASF to the global pig industry.

Here we discuss ASFV infection of domestic pigs, wild boar and other wildlife hosts, summarizing current knowledge of how host and viral genetics contribute to pathogenesis and the different disease outcomes seen in different hosts. We also discuss prospects of how these differences might be leveraged to inform breeding or genetic engineering strategies to improve disease resistance in the domestic pig population.

ASFV GENETICS

ASFV Genetic Variability

African swine fever virus is a large double-stranded DNA virus, which replicates predominantly in the cell cytoplasm and shares a similar replication cycle and genome structure with the poxviruses. However the icosahedral virus morphology differs from the poxviruses and genome sequencing established that ASFV is the only member of a unique virus family, the *Asfarviridae*. Genome sequencing also showed a distant relationship between ASFV and some giant viruses that infect lower eukaryotes, including the Faustovirus, Pacmanvirus and

Kamobaviruses (Reteno et al., 2015; Bajrai et al., 2016; Andreani et al., 2017). Thus these diverse viruses may have shared a common ancestor. ASFV's genome varies between 170 and 190 kb in length. These gross differences in genome size are predominately due to differences in the copy number of five different multigene families (MGF); for example the copy number of MGF 360 can vary between 11 and 18 in field isolates (Chapman et al., 2008). Promotion of homologous recombination or unequal crossover during DNA replication within infected cells (Rodríguez et al., 1992) is a likely driver of the loss and exchange of genetic material that has been observed in isolates from both ticks and domestic pigs (Dixon and Wilkinson, 1988; Chapman et al., 2011). Interestingly the rapid amplification of individual genes by gene duplication under selection pressure has been observed in poxviruses (Elde et al., 2012) and a similar mechanism may have contributed to differences in the copy number of individual MGFs in ASFV. Paralogs of the MGF genes can be very divergent in sequence indicating evolution over an extended period. This may be related to selection pressure exerted such as altered host tropism. Gene families have evolved in other viruses, for example vaccinia virus encodes a highly divergent family of proteins containing a Bcl-2 protein fold which have different roles in evasion of innate immune responses, including apoptosis and signaling pathways (Graham et al., 2008; Kvensakul et al., 2008; Neidel et al., 2015). Errors in unit genome resolution during the head-to-head DNA replication can also result in sequence transpositions from one genome end to the other as seen in a recent ASFV isolate from northern Estonia that had 14 kb deleted from the left end of the genome and replaced with 7 kb from the right (Zani et al., 2018).

ASFV isolates have been divided into genotypes based on partial sequencing of the B646L gene which encodes the ASFV major capsid protein. This has defined 24 different genotypes to date (Figure 2; Bastos et al., 2003; Quembo et al., 2018) which fall into three main lineages (Boshoff et al., 2007). However, a limitation of this approach is that the number of nucleotide differences between closely related genotypes can be low. Nevertheless phylogenetic trees constructed from short stretches of conserved genes such as p72 do broadly fit with those generated from concatenated conserved nucleotide or protein sequences (de Villiers et al., 2010). Currently 30 complete genome sequences are available but more than two-thirds of these are of three related genotypes and are therefore clearly not representative of the 24 described genotypes based on p72 sequencing, limiting opportunities to infer evolutionary relationships. Up to 18 genes under positive selection for diversification have been identified by comparing rates of synonymous to non-synonymous substitutions at individual amino acids (de Villiers et al., 2010). These included members of MGF 360 and 505 families, genes involved in modulating host cell functions, several enzymes, the CD2-like and C-type lectin genes and the virus capsid protein chaperone B602L. Drivers for diversification might include immune or host genetic pressure. The major capsid protein did not have any sites under strong selection indicating strong stabilizing selection (de Villiers et al., 2010).

ASFV Modulation of the Host Response in the Domestic Pig

Inhibitors of Type I Interferon

ASFV encodes a number of proteins that inhibit innate immune responses including type I interferon (IFN), the main antiviral response. Stimulation of cellular pattern recognition receptors by an array of pathogen associated molecular patterns induces signaling pathways leading to transcription of type I IFN (Bowie and Unterholzner, 2008; Schoggins et al., 2011; Thompson et al., 2011). The secreted type I IFN activates signaling in infected and bystander cells leading to transcription of over 300 interferon stimulated genes. These include proteins that induce an antiviral state, via blocking the viral replication cycle or activating components of protective innate and adaptive immune responses (Schoggins et al., 2011). For example Mx proteins sequester viral replication factors preventing efficient replication (Netherton et al., 2009) and IFITM proteins restrict virus entry by inhibiting membrane fusion (Benfield et al., 2015).

Although their functional role is currently poorly understood, and they have no obvious similarity to other genes or proteins, there is mounting evidence to suggest that MGF genes may play a role in both host range and subversion of the innate immune system. Sequence analysis indicated that the low virulence isolate OUR T88/3 lacks eight MGF genes (MGF360-10L, 11L, 12L, 13L, 14L, MGF505-1R, 2R, 3R), which are otherwise present in virulent ASFV isolates, suggesting they may play a role in virulence (Chapman et al., 2008; Dixon et al., 2013). Furthermore, levels of IFN in the bloodstream apex prior to the viremic peak, indicating the ability of virulent viruses to endure the host IFN response (Karalyan et al., 2012; Golding et al., 2016). Indeed IFN priming of primary macrophages limited replication of attenuated OURT88/3 but not virulent isolates (Golding et al., 2016).

ASFV lacking these specific MGF genes, including genetically modified virus with the genes in question deleted, induce a stronger innate immune response. Deletion of five MGF360 and three MGF505 from highly virulent Benin 97/1 resulted in attenuation, increased IFN β production *in vitro* and significantly enhanced protection *in vivo* against challenge with parental virus (Reis et al., 2016). The presence of genes from the MGF360 and MGF505 cluster are directly responsible for suppressing IFN responses *in vitro* in cells infected with virulent Pr4 (Afonso et al., 2004) and overcoming IFN mediated inhibition of virus replication (Golding et al., 2016). Further experiments were not able to directly attribute this function to a sole gene or MGF family. A subset of these genes are also important for host range in *Ornithodoros* ticks (Burrage et al., 2004), however the mode of action in the arthropod vector is unknown. Other ASFV genes shown to inhibit type I IFN responses include I329L, an agonist of Toll-like receptor 3 signaling (de Oliveira et al., 2011).

Inhibitors of Apoptosis

Induction of apoptosis can limit virus replication and many viruses, including ASFV, encode apoptosis inhibitors (Dixon et al., 2017). These include a Bcl-2 family member A179L, inhibitor of apoptosis member A224L and a C-type lectin protein

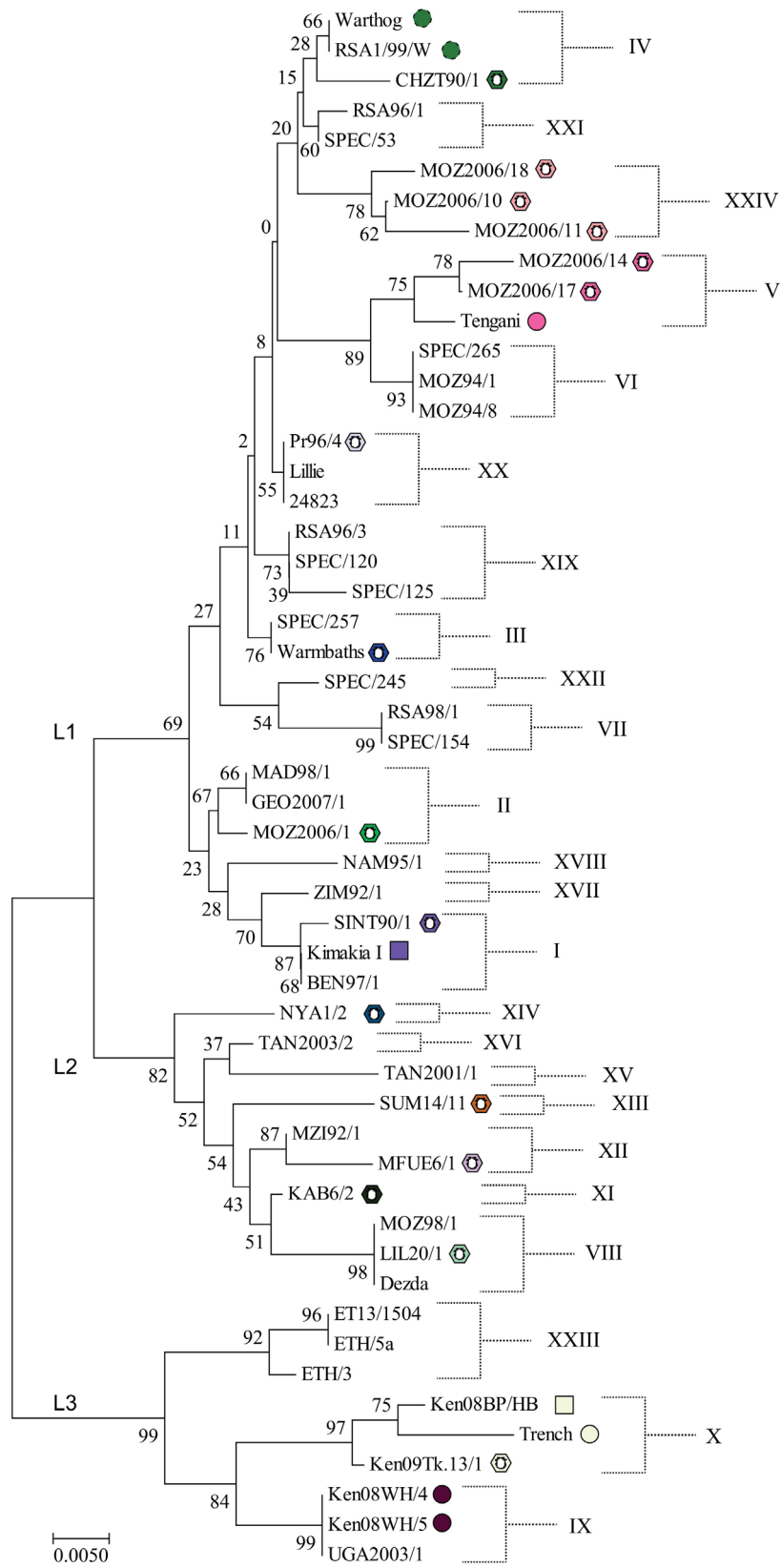


FIGURE 2 | Continued

FIGURE 2 | Neighbor-Joining phylogenetic tree of representative ASFV isolates. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.29203136 is shown. The bootstrap test values (i.e., percentage of replicate trees in which the associated taxa clustered together, 1000 replicates) are shown next to the nodes. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. The analysis involved 47 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 399 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). Symbols indicate isolates shown in **Figure 3**, all other isolates were obtained from domestic pigs. L1, L2, and L3 indicate the lineages identified by Boshoff et al. (2007). Full details of these isolates are provided in **Supplementary Table S1**.

EP153R (Hurtado et al., 2004). Other ASFV proteins inhibit stress-activated apoptosis (Zhang et al., 2010). The A179L protein has an unusually broad specificity of binding to pro-apoptotic Bcl-2 family BH3 domain-containing proteins (Banjara et al., 2017). This may allow for functionality in both mammalian and *Ornithodoros* hosts.

Adhesion Proteins

The ASFV CD2-like protein causes binding of red blood cells to extracellular virions and infected cells. This protein has roles in virus dissemination and persistence in blood in the mammalian host (Borca et al., 1998) as well as facilitating virus uptake into the tick vector (Rowlands et al., 2009). Both of these functions may provide an advantage for virus replication in the tick-warthog cycle.

INFECTION AND PATHOGENESIS IN DIFFERENT HOST SPECIES

ASFV in Domestic Pigs

Transmission of ASFV to Domestic Pigs and Wild Boar

The emergence of ASFV from its ancient sylvatic ecology in Eastern and Southern Africa, involving warthogs and soft tick vectors of the *Ornithodoros* spp., into domestic pigs and wild boar has resulted in a dramatic change in the pathogenesis of the virus and the mechanisms by which transmission occurs. Transmission by the tick vector is not required in the domestic pig or wild boar cycle and direct transmission between pigs occurs readily in the absence of the tick vector (epidemiological cycles 3 and 4 in **Figure 1**; Wilkinson, 1984; Guinat et al., 2016). Indeed the ancient sylvatic cycle involving warthogs and ticks has only been described in parts of Eastern and Southern Africa (Jori et al., 2013) meaning that spread through other susceptible populations is unlikely to have placed the same constraints on virus replication.

ASF first spread outside Africa, to Portugal and Spain and from there to a number of other European countries, as well as Brazil and the Caribbean. The disease persisted in the Iberian Peninsula for over three decades, but was eradicated from all of these countries except Sardinia by the mid-1990s. In the Iberian Peninsula ASFV circulated in pigs, wild boar and *Ornithodoros erraticus*, whereas soft ticks did not play a role in Sardinia (Sánchez-Vizcaíno et al., 2015). However, wild boar were not thought to play a significant role in maintaining the virus. In the present epidemic in Russia and Eastern Europe, wild boar have played an important role in spread of disease and maintaining a

wildlife reservoir (Abrahantes et al., 2017; Chenais et al., 2018) and there is no evidence of a role for soft ticks. Wild boar show similar clinical signs to domestic pigs and case fatality rates are also close to 100% following infection with highly virulent isolates (Gabriel et al., 2011). Studying the evolution of ASFV clinical forms and associated viral genetic changes during the current epidemic in Europe provides an excellent opportunity to follow the virus adaptation to different hosts.

Pathogenesis of ASFV in Domestic Pigs and Wild Boar

Early descriptions of ASF disease in domestic pigs were of an acute hemorrhagic fever causing death of close to 100% of infected pigs (Montgomery, 1921). This is still the predominant disease form reported in both Africa and in Europe (Tauscher et al., 2015). However different disease courses in pigs have been associated with isolates which vary in virulence. Moderately virulent isolates result in death of a lower percentage of animals and a subacute form of the disease. Low virulence isolates may cause few if any deaths and a chronic form of disease characterized by the absence of vascular lesions but signs such as delayed growth, emaciation, joint swelling, skin ulcers and lesions associated with secondary bacterial infection. Moderately virulent and low virulence isolates were described after the introduction of the virus into Spain and Portugal, and similar isolates have now been described from different countries in Africa (Souto et al., 2016) and also in Eastern Europe (Gallardo et al., 2018; Zani et al., 2018). Detection of ASFV specific antibodies in serum from wild boar in Eastern Europe may indicate reduced virulence of circulating isolates, since in acute infections animals die before an antibody response is detected. As yet limited full genome sequences are available for ASFV but reduction in virulence has been associated with genome changes including large deletions and sequence transpositions from one genome end to the other (Zani et al., 2018). Recovered animals may remain persistently infected over extended time periods of weeks or months. Shedding of virus and transmission from recovered animals to in contact animals has been described but it remains unclear whether these carrier animals play an important role in virus spread (Boinas et al., 2004; de Carvalho Ferreira et al., 2012; Gallardo et al., 2015; Petrov et al., 2018). Interestingly ASFV has persisted in Sardinia for 40 years in a pig-pig-wild boar transmission cycle without loss of virulence and with few genetic changes (Granberg et al., 2016; Sanna et al., 2017). Therefore, the main mechanisms of ASFV persistence and transmission in different epidemiological scenarios clearly influence which types of ASFV isolates emerge and become predominant.

Influence of Host Genetics on the Outcome of Disease in Pigs

Most reports of ASF disease in domestic pigs or wild boar describe similar acute disease forms with high case fatality in all ages and breeds following infection with highly virulent ASFV isolates (Gabriel et al., 2011; Blome et al., 2013; Nurmoja et al., 2017). However there are also reports indicating differences in susceptibility to disease in some populations or ages of domestic pigs. In one study the percentage of older pigs surviving infection with a moderately virulent isolate was shown to be higher than for younger pigs (Post et al., 2017). In Mozambique, although some pigs were identified that survived infection with a virulent isolate, this apparent resistance was found not to be transmitted to offspring based on results of viral challenge experiments (Penrith et al., 2004). In some regions of Africa apparently healthy pigs have tested positive for virus or had ASFV specific antibodies without showing clinical signs of the disease (Uttenthal et al., 2013; Thomas et al., 2016; Abworo et al., 2017; Kipanyula and Nong'ona, 2017). SNP analysis was used to assess the genetic diversity of two populations of Kenyan pigs and compare them to bushpigs, warthogs, European wild boar as well as four breeds of commercial pigs. Principal component and admixture analyses identified six separate groups, with the two populations of Kenyan pigs forming two distinct groups alongside groups comprised of wild boar, Duroc pigs, African suids or the three other domestic pig breeds (Large white cross, Yorkshire, and Landrace). The failure to resolve bushpigs and warthogs as separate populations was likely due to few markers in the porcine SNP array being amplified in samples from these animals. The Homabay population from Kenya had a local indigenous composition distinct from commercial breeds. In contrast, pigs from Busia and the surrounding area were a non-homogenous admixed population with significant introgression of genes from commercial breeds. Notably a higher percentage of pigs that tested negative for ASFV by PCR had significantly higher local ancestry. Although serology was not performed to prove previous ASFV infection, the study provides some evidence that local ancestry confers a survival advantage against ASFV and a basis to explore genetic determinants underlying resistance to developing disease (Mujibi et al., 2018).

ASFV in Other Suid Species

The link between outbreaks of ASF and a wildlife reservoir was suspected during the emergence of the disease in the early twentieth century (Montgomery, 1921). Subsequent studies confirmed the isolation of infectious virus from apparently healthy warthogs associated with outbreaks of disease in domestic pigs in both Kenya and South Africa (De Kock et al., 1940; Hammond and Detray, 1955). Infectious virus has been recovered from bushpigs (*Potamochoerus* spp.), warthogs (*Phacochoerus* spp.), *Ornithodoros* ticks and a single giant forest hog (*Hylochoerus meinertzhageni*). The expansion of ASF into South-East Asia raises the possibility of transmission of the virus to other species and genera of suids which have not previously encountered the disease. Warty pigs and bearded pigs (all species of *Sus*) indigenous to Indonesia and the Philippines

would be predicted to suffer similar disease outcomes to domestic pigs and wild boar. However, pygmy hogs (*Porcula salvania*) found in India and babirusa (*Babirusa* spp.) from Indonesia are distinct genera (Funk et al., 2007) and their susceptibility to ASFV is unclear; although classical swine fever virus, an RNA virus that causes a disease with similar clinical signs to ASFV, can infect and kill pygmy hogs (Barman et al., 2012). The wild populations of many of these species are of concern with pygmy hogs and Visayan warty pigs (*Sus cebifrons*) considered critically endangered according to the International Union for Conservation of Nature (Narayan et al., 2008; Meijaard et al., 2017). Spill over of ASF into these wild suids could lead to other avenues for exploring disease resistance, but could add an unwelcome pressure on already threatened populations.

ASFV in *Potamochoerus* spp.

Bushpigs (*Potamochoerus larvatus*) are distributed throughout Eastern and Southern Africa while red river hogs (*Potamochoerus porcus*) are found in sub-Saharan West and Central Africa. ASFV has been isolated from both bushpigs and red river hogs (Figure 3) and as the two species are closely related and can interbreed we will use bushpigs to refer to all *Potamochoerus* spp. ASFV infection does not induce clinical signs of disease in bushpigs, with virus titres in the blood and tissues 100-fold lower than the 8–9 logs typically seen in domestic pigs (Anderson et al., 1998; Oura et al., 1998a). Virus replication in tissues is also reduced and although extensive B-cell apoptosis in lymph nodes has been observed, this is not as extensive as seen in domestic pigs and other structures are essentially unaffected. Experimentally infected bushpigs clear ASFV from the tissues (Detray, 1963; Anderson et al., 1998) and gain immunity to subsequent rechallenge with homologous virus strains. Bushpigs can transmit virus to feeding ticks and to in-contact pigs. Transmission to pigs depends on the frequency of contacts with domestic pigs and may also be strain specific (Anderson et al., 1998). The role of bushpigs in maintaining a reservoir of virus is unclear since they do not reside in burrows like warthogs and hence are not thought to come into frequent contact with *Ornithodoros* ticks.

ASFV in *Phacochoerus* spp.

There are two species of warthog in Africa, the common warthog (*Phacochoerus africanus*) which is distributed throughout sub-Saharan Africa and the desert warthog (*Phacochoerus aethiopicus*) which is restricted to the Horn of Africa and Northern Kenya. We will use the term warthog to refer to the common warthog as to our knowledge ASFV has not been isolated from *Phacochoerus* spp. within the known distribution of the desert warthog, although the recent outbreaks in the Tigray region in northern Ethiopia (Achenbach et al., 2017) suggest that it may only be a matter of time before this occurs. ASFV has been isolated from warthogs across Southern and Eastern Africa (Figure 1), and seropositive animals have also been found in Botswana and Zimbabwe (Jori et al., 2013).

Serological surveys suggest that infection rates in populations of warthogs where ASFV is endemic are typically greater than 80% (Plowright, 1977; Thomson, 1985), although viremia in wild

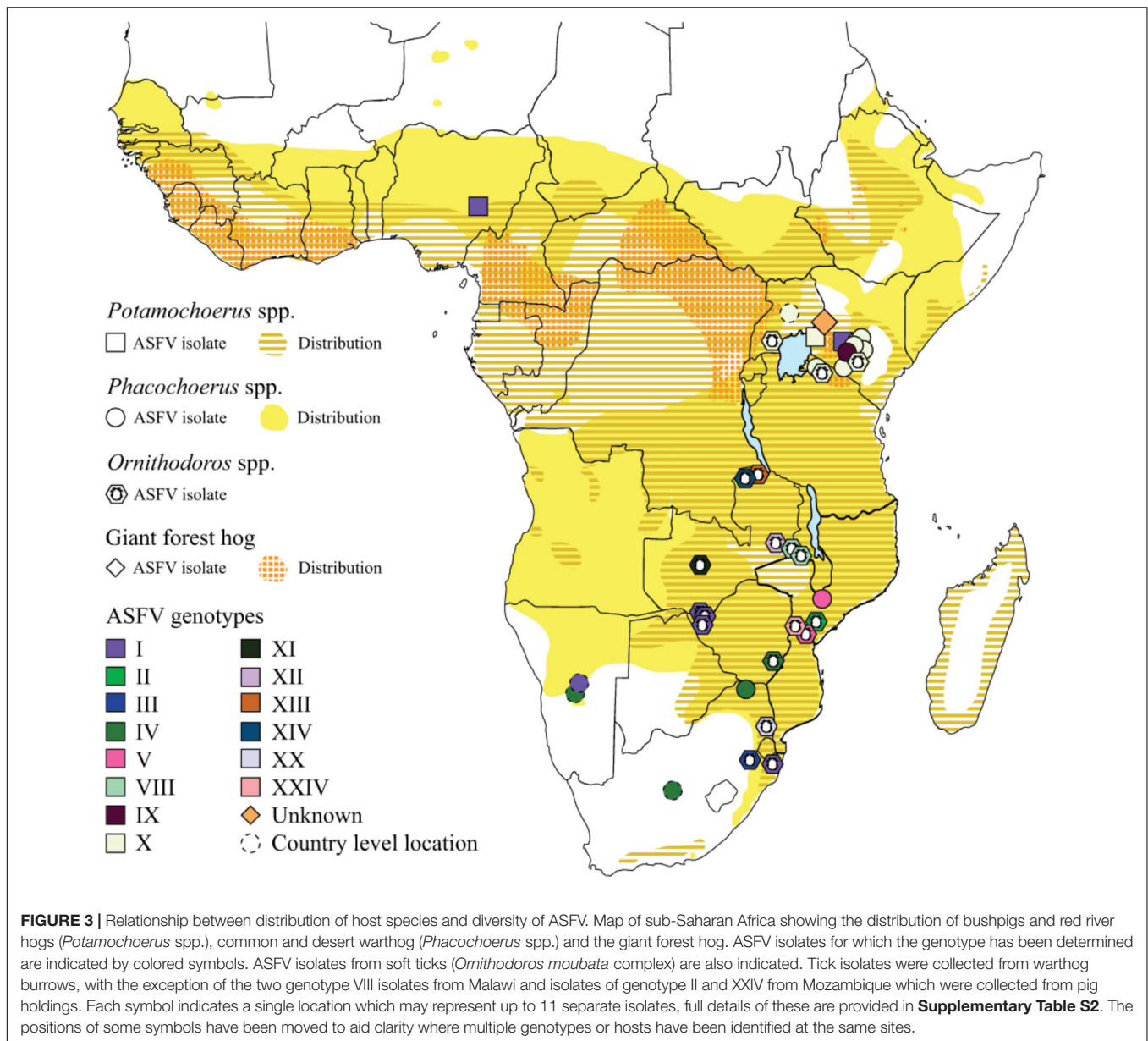


FIGURE 3 | Relationship between distribution of host species and diversity of ASFV. Map of sub-Saharan Africa showing the distribution of bushpigs and red river hogs (*Potamochoerus* spp.), common and desert warthog (*Phacochoerus* spp.) and the giant forest hog. ASFV isolates for which the genotype has been determined are indicated by colored symbols. ASFV isolates from soft ticks (*Ornithodoros moubata* complex) are also indicated. Tick isolates were collected from warthog burrows, with the exception of the two genotype VIII isolates from Malawi and isolates of genotype II and XXIV from Mozambique which were collected from pig holdings. Each symbol indicates a single location which may represent up to 11 separate isolates, full details of these are provided in **Supplementary Table S2**. The positions of some symbols have been moved to aid clarity where multiple genotypes or hosts have been identified at the same sites.

adult warthogs is rare, with infectious virus mostly restricted to lymph nodes. However, wild caught neonatal animals from the Serengeti do exhibit detectable viremia (Plowright, 1977; Thomson, 1985) and experimental infection of naïve young warthogs also yields low viremia for several weeks which may be sufficient to infect ticks (Thomson et al., 1980; Anderson et al., 1998). Therefore it is likely that the warthog-tick sylvatic cycle is in part maintained by ticks transmitting the disease to 3–4 weeks old warthogs that can then transmit the virus to naïve ticks. Interestingly the proportion of ASFV positive ticks in warthog burrows in Western Uganda were found to be very low and the majority of warthogs in this area did not become seropositive until they were 6 months old (Plowright, 1977). In combination with the observation that warthogs in the central Kenyan highlands were seropositive in the absence of ticks,

this suggests there a number of different sylvatic cycles capable of maintaining a virus reservoir. Infectious virus persisted in warthog tissues up to 25 weeks post-infection, but is cleared by 56 weeks (Anderson et al., 1998). Field observations have demonstrated persistent infection of warthog tissues (Plowright et al., 1969b; Plowright, 1977). This could be explained by repeated re-infection of warthogs by ticks with the same virus strain. Warthogs probably develop an adaptive immune response to a given ASFV strain, which while insufficient to prevent replication at primary sites of infection can prevent an acute phase and hence virus dissemination into the blood stream.

ASFV in *Ornithodoros* spp.

African swine fever has been isolated from *Ornithodoros* spp. ticks collected from warthog burrows from Kenya to South Africa

(see **Figure 1**), although the proportion of ticks positive for virus is typically less than 1%. Virus is transmitted sexually and transtadially in ticks (Plowright et al., 1970) and can be isolated from all developmental stages (Plowright et al., 1969a; Quembo et al., 2018). Transovarial transmission of the virus has also been shown in ticks from the *O. moubata* complex. Detailed genetic and morphological analyses of Afrotropical *Ornithodoros* spp. have identified at least four species within each of the *O. savignyi* and *O. moubata* groups, only one of which is not thought to be associated with pigs or warthogs (Bakkes et al., 2018). However, *O. phacochoerus*, *O. porcinus*, and *O. waterbergensis* are the principal species linked to the sylvatic cycle. Although *O. moubata* spp. are true biological vectors of ASFV virus replication can be deleterious to the tick (Hess et al., 1989) and experimental infection of *Ornithodoros* spp. from the Americas also causes tick mortality (Hess et al., 1987). Therefore, the relative ability of different Afrotropical *Ornithodoros* species to support the replication of different strains of ASFV may be an important aspect of the sylvatic cycle. Genetically related, but distinct strains of ASFV have been identified in ticks from separate warthog burrows within close proximity to each other (Dixon and Wilkinson, 1988; Wilkinson et al., 1988), demonstrating divergent evolution of ASFV within the sylvatic cycle. The sylvatic cycle in Africa provides a reservoir of persistently infected hosts to maintain the infection. In the current situation in Europe and China wild boar and domestic pigs in most cases develop disease with high levels of case fatality. Thus maintaining a virus reservoir requires a readily available pool of susceptible hosts or an effective indirect transmission route.

As yet few virus genes have been identified which confer an advantage for replication in the tick vector (Burrage et al., 2004; Rowlands et al., 2009). A functional genomics approach, involving targeted gene deletions and modifications and testing the effect of these on virus replication in the tick would provide further insights. The lack of a tick cell line susceptible to ASFV infection is a constraint meaning that infections of live ticks is required to achieve this. Further comparative full genome sequencing of virus isolates from tick/warthog and domestic-pig/wild-boar cycles would also help to unravel virus adaptations and selections required for replication in the tick.

POTENTIAL MECHANISMS FOR HOST RESISTANCE

Due to the paucity of experimental and genetic data available it is difficult to draw conclusions about why warthogs and bushpigs exhibit limited clinical signs after infection with ASFV when compared to domestic pigs and wild boar. Viral replication is approximately 100-fold lower in bushpigs than in domestic pigs, and replication in warthogs 10-fold less than bushpigs. Comparison of *in vitro* growth curves in macrophages suggest there is no intrinsic difference in the ability of target cells to support the growth of ASFV between the three species (Anderson et al., 1998). It is therefore likely that the innate immune response plays a key role in controlling the levels of virus replication and

pathogenesis in different infected hosts. Thus in hosts which do not develop disease the innate immune response may better control virus replication and avoid a pathogenic response. This may involve both viral and host factors. For example virus genetic factors may be less effective in controlling innate responses in the wild African suids compared to the domestic pig or wild boar. Alternatively host genetic factors may reduce over-activation of potentially harmful responses and hence reduced clinical signs may also be due to host tolerance.

ASFV encodes for a diverse combination of genes capable of suppressing the induction of type I IFN in domestic pigs. It is tempting to speculate that this functional redundancy of viral IFN inhibitory factors evolved to combat the effect of IFN in the natural host. It would therefore be interesting to compare type I IFN induction and responses in wild African suids compared to domestic pigs and wild boar. Human IFN stimulated genes Mx1 and IFITM (Netherton et al., 2009; Munoz-Moreno et al., 2016) inhibit ASFV replication *in vitro*, however the effect of the *suid* homologs are unknown. Work in our laboratories is currently ongoing to determine the genetic and functional differences between the pig and warthog homologs of these genes.

NK cells are capable of killing virus infected cells, secreting immunomodulatory cytokines and activating dendritic cells, linking with the adaptive immune response. Subclinical infections of domestic pigs with low virulent strains of ASFV and protection in subsequent challenge studies are linked to enhanced NK cell activity (Leitao et al., 2001), so differences in the way these cells respond to ASFV could play a role in the ability of bushpigs and warthogs to control infection.

Interspecies differences in the pathology of ASFV could also be linked to differences in host response to infection. Like many hemorrhagic diseases the pathology of ASFV in domestic pigs has been linked to the overexpression of cytokines such as IFN and tumor necrosis factor alpha (Oura et al., 1998b; Gómez del Moral et al., 1999; Golding et al., 2016). The NF- κ B transcription factor controls transcription of both these cytokines and ASFV encodes proteins that can inhibit this pathway (Granja et al., 2006), but these viral NF- κ B inhibitors could be less effective in the warthogs and bushpigs compared to domestic pigs or wild boar. Alternatively host transcription factors may be less active in warthogs and bushpigs. For example reporter assays in monkey kidney cells and mouse embryonic fibroblasts show that the RELA subunit of NF- κ B from the domestic pig has lower activity after induction by external stimuli than the warthog homolog, but has higher basal activity (Palgrave et al., 2011) and that this difference appears to be due a S531P variant present in the warthog. Genome sequences will help develop additional avenues of research to understand the mechanisms responsible for differences in disease outcomes between domestic pigs, bushpigs and warthogs.

CONCLUSION AND FUTURE PERSPECTIVES

The mechanisms which result in reduced viral replication and lack of disease in African wild suids after ASFV infection are

largely unknown. The data so far indicate that this is not due to an intrinsic difference in the ability of the virus to replicate in macrophages from these hosts. A more likely explanation is that the innate immune system of these hosts is better able to control virus replication resulting in a reduced systemic infection and reduced pathogenesis. This may involve a balance between virus and host factors which has evolved over long term infections of these hosts. Sequence information from African wild suids will enable further investigation of the interaction of ASFV with components of the innate immune system compared to domestic pigs and wild boar. A better understanding of ASFV mechanisms of evading host defenses will contribute to this. Of special interest are the functions of the many members of five MGF encoded by ASFV. As is the case in other viruses these may have evolved in the virus genome to modulate the host's innate immune response.

Genetic modification has been used to generate pigs resistant to porcine respiratory and reproductive syndrome virus (Whitworth et al., 2016; Burkard et al., 2018) or classical swine fever virus (Xie et al., 2018) and therefore could be a viable route to increase resistance to ASFV. Identified warthog or bushpig sequences could be engineered into the pig genome to generate animals in which replication and/or disease burden after ASFV infection is reduced. However, in order to generate a pig that is fully resistant to ASFV infection, as has been accomplished with porcine respiratory and reproductive syndrome virus, a more effective strategy may be to target essential elements of the viral replication cycle such as entry.

Different clinical courses of ASFV infection in pigs have been described, apparently largely due to the virulence of the virus isolates, and sequencing the genomes of isolates of reduced virulence have identified virus genes associated with this phenotype. Targeted gene modifications and deletions and testing of the genetically modified viruses in macrophages and in pigs have contributed to understanding of virulence factors and how the virus modulates host responses. There are no licensed ASFV vaccines available and further research in this area will also contribute to the development of live attenuated vaccines for ASFV.

The issue of whether outcome of ASFV infection in pigs also depends on host genetics has been discussed and considered over a number of years without definite conclusions. Recent studies linking genetics of different pig breeds in Kenya with prevalence

of ASFV infection is a promising step forward. Further study of these pigs to confirm that resistance to developing disease after ASFV infection is due to genetic differences rather than hitherto unknown environmental factors could open the possibility of breeding in resistance to the disease. Analysis of other African pig breeds with suspected disease resistance to ASFV may identify additional factors that could be incorporated into such a strategy. Viable bushpig-domestic pig hybrids have been observed in the field and these could open up another avenue of research if these animals were also resistant to disease.

In the longer term a better understanding of ASFV interactions with its different hosts will be not only of great scientific interest but will lead to improved control strategies for this disease and help prevent global spread.

AUTHOR CONTRIBUTIONS

CN, SC, CB, and LD wrote individual sections of the manuscript. All authors contributed to the editing and revision of the manuscript as well as read and approved the submitted version.

FUNDING

CN and LD were supported by the United Kingdom Biotechnology and Biological Sciences Research Council (BBSRC) through projects BBS/E/I/00007030, BBS/E/I/00007031, BBS/E/I/00007034, BBS/E/I/00007034, BBS/E/I/00007035, BBS/E/I/00007036, BBS/E/I/00007037, BBS/E/I/00007038, and BBS/E/I/00007039, and by the United Kingdom Department for Environment, Food & Rural Affairs through project SE1516. SC was supported by The Oxford Interdisciplinary Bioscience Doctoral Training Partnership funded by the BBSRC grant BB/M011224/1. CB was supported by a Royal Veterinary College Internal Grant (IGS 2930).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00402/full#supplementary-material>

REFERENCES

- Abrahantes, J. C., Gogin, A., Richardson, J., and Gervelmeyer, A. (2017). Epidemiological analyses on African swine fever in the Baltic countries and Poland. *EFSA J.* 15:e04732. doi: 10.2903/j.efsa.2017.4732
- Abworo, E. O., Onzere, C., Oluoch Amimo, J., Riitho, V., Mwangi, W., Davies, J., et al. (2017). Detection of African swine fever virus in the tissues of asymptomatic pigs in smallholder farming systems along the Kenya-Uganda border: implications for transmission in endemic areas and ASF surveillance in East Africa. *J. Gen. Virol.* 98, 1806–1814. doi: 10.1099/jgv.0.000848
- Achenbach, J. E., Gallardo, C., Nieto-Pelegrin, E., Rivera-Arroyo, B., Degefa-Negi, T., Arias, M., et al. (2017). Identification of a new genotype of african swine fever virus in domestic pigs from Ethiopia. *Transbound. Emerg. Dis.* 64, 1393–1404. doi: 10.1111/tbed.12511
- Afonso, C. L., Piccone, M. E., Zaffuto, K. M., Neilan, J., Kutish, G. F., Lu, Z., et al. (2004). African swine fever virus multigene family 360 and 530 genes affect host interferon response. *J. Virol.* 78, 1858–1864. doi: 10.1128/jvi.78.4.1858-1864.2004
- Anderson, E. C., Hutchings, G. H., Mukarati, N., and Wilkinson, P. J. (1998). African swine fever virus infection of the bushpig (*Potamochoerus porcus*) and its significance in the epidemiology of the disease. *Vet. Microbiol.* 62, 1–15. doi: 10.1016/s0378-1135(98)00187-4
- Andreani, J., Khalil, J. Y. B., Sevvana, M., Benamar, S., Di Pinto, F., Bitam, I., et al. (2017). Pacmanvirus, a new giant icosahedral virus at the crossroads between Asfarviridae and Faustoviruses. *J. Virol.* 91:e00212-17. doi: 10.1128/JVI.00212-17

- Bajrai, L. H., Benamar, S., Azhar, E. I., Robert, C., Levasseur, A., Raoult, D., et al. (2016). Kaumobavirus, a new virus that clusters with Faustoviruses and Asfarviridae. *Viruses* 8:E278. doi: 10.3390/v8110278
- Bakkes, D. K., De Klerk, D., Latif, A. A., and Mans, B. J. (2018). Integrative taxonomy of afrotropical ornithodoros (Ornithodoros) (Acari: Ixodida: Argasidae). *Ticks Tick Borne Dis.* 9, 1006–1037. doi: 10.1016/j.ttbdis.2018.03.024
- Banjara, S., Caria, S., Dixon, L. K., Hinds, M. G., and Kvensakul, M. (2017). Structural insight into african swine fever virus A179L-mediated inhibition of apoptosis. *J. Virol.* 91:e02228–16. doi: 10.1128/JVI.02228-16
- Barman, N. N., Bora, D. P., Tiwari, A. K., Kataria, R. S., Desai, G. S., and Deka, P. J. (2012). Classical swine fever in the pygmy hog. *Rev. Sci. Tech.* 31, 919–930.
- Bastos, A. D., Penrith, M. L., Cruciare, C., Edrich, J. L., Hutchings, G., Roger, F., et al. (2003). Genotyping field strains of African swine fever virus by partial p72 gene characterisation. *Arch. Virol.* 148, 693–706. doi: 10.1007/s00705-002-0946-8
- Benfield, C. T. O., Smith, S. E., Wright, E., Walsh, R. S., Ferrara, F., Temperton, N. J., et al. (2015). Bat and pig IFN-induced transmembrane protein 3 restrict cell entry by influenza virus and lyssaviruses. *J. Gen. Virol.* 96, 991–1005. doi: 10.1099/vir.0.000058
- Blome, S., Gabriel, C., and Beer, M. (2013). Pathogenesis of African swine fever in domestic pigs and European wild boar. *Virus Res.* 173, 122–130. doi: 10.1016/j.virusres.2012.10.026
- Boinas, F. S., Hutchings, G. H., Dixon, L. K., and Wilkinson, P. J. (2004). Characterization of pathogenic and non-pathogenic African swine fever virus isolates from Ornithodoros erraticus inhabiting pig premises in Portugal. *J. Gen. Virol.* 85(Pt 8), 2177–2187. doi: 10.1099/vir.0.80058-0
- Borca, M. V., Carrillo, C., Zsak, L., Laegreid, W. W., Kutish, G. F., Neilan, J. G., et al. (1998). Deletion of a CD2-like gene, 8-DR, from African swine fever virus affects viral infection in domestic swine. *J. Virol.* 72, 2881–2889.
- Boshoff, C. I., Bastos, A. D., Gerber, L. J., and Vosloo, W. (2007). Genetic characterisation of African swine fever viruses from outbreaks in southern Africa (1973–1999). *Vet. Microbiol.* 121, 45–55. doi: 10.1016/j.vetmic.2006.11.007
- Bowie, A. G., and Unterholzner, L. (2008). Viral evasion and subversion of pattern-recognition receptor signalling. *Nat. Rev. Immunol.* 8, 911–922. doi: 10.1038/nri2436
- Burkard, C., Opriessnig, T., Mileham, A. J., Stadejek, T., Ait-Ali, T., Lillico, S. G., et al. (2018). Pigs lacking the scavenger receptor cysteine-rich domain 5 of CD163 are resistant to porcine reproductive and respiratory syndrome virus 1 infection. *J. Virol.* 92:e00415–18. doi: 10.1128/JVI.00415-18
- Burrage, T. G., Lu, Z., Neilan, J. G., Rock, D. L., and Zsak, L. (2004). African swine fever virus multigene family 360 genes affect virus replication and generalization of infection in Ornithodoros porcinus ticks. *J. Virol.* 78, 2445–2453. doi: 10.1128/jvi.78.5.2445-2453.2004
- Chapman, D. A., Darby, A. C., Da Silva, M., Upton, C., Radford, A. D., and Dixon, L. K. (2011). Genomic analysis of highly virulent Georgia 2007/1 isolate of African swine fever virus. *Emerg. Infect. Dis.* 17, 599–605. doi: 10.3201/eid1704.101283
- Chapman, D. A., Tcherepanov, V., Upton, C., and Dixon, L. K. (2008). Comparison of the genome sequences of non-pathogenic and pathogenic African swine fever virus isolates. *J. Gen. Virol.* 89(Pt 2), 397–408. doi: 10.1099/vir.0.83343-0
- Chenais, E., Ståhl, K., Guberti, V., and Depner, K. (2018). Identification of wild boar-habitat epidemiologic cycle in African swine fever epizootic. *Emerg. Infect. Dis.* 24, 810–812. doi: 10.3201/eid2404.172127
- de Carvalho Ferreira, H. C., Weesendorp, E., Elbers, A. R., Bouma, A., Quak, S., Stegeman, J. A., et al. (2012). African swine fever virus excretion patterns in persistently infected animals: a quantitative approach. *Vet. Microbiol.* 160, 327–340. doi: 10.1016/j.vetmic.2012.06.025
- De Kock, G., Robinson, E. M., Keppel, J. J. G., and DuToit, P. J. (1940). Swine fever in South Africa. *Onderstepoort J. Vet. Res. Anim. Ind.* 14, 31–93.
- de Oliveira, V. L., Almeida, S. C., Soares, H. R., Crespo, A., Marshall-Clarke, S., and Parkhouse, R. M. (2011). A novel TLR3 inhibitor encoded by African swine fever virus (ASFV). *Arch. Virol.* 156, 597–609. doi: 10.1007/s00705-010-0894-7
- de Villiers, E. P., Gallardo, C., Arias, M., da Silva, M., Upton, C., Martin, R., et al. (2010). Phylogenomic analysis of 11 complete African swine fever virus genome sequences. *Virology* 400, 128–136. doi: 10.1016/j.virol.2010.01.019
- Detray, D. E. (1963). African swine fever. *Adv. Vet. Sci.* 8, 299–333.
- Dixon, L. K., Chapman, D. A., Netherton, C. L., and Upton, C. (2013). African swine fever virus replication and genomics. *Virus Res.* 173, 3–14. doi: 10.1016/j.virusres.2012.10.020
- Dixon, L. K., Sanchez-Cordon, P. J., Galindo, I., and Alonso, C. (2017). Investigations of Pro- and anti-apoptotic factors affecting african swine fever virus replication and pathogenesis. *Viruses* 9:E249. doi: 10.3390/v9090241
- Dixon, L. K., and Wilkinson, P. J. (1988). Genetic diversity of African swine fever virus isolates from soft ticks (*ornithodoros moubata*) inhabiting warthog burrows in Zambia. *J. Gen. Virol.* 69, 2981–2993. doi: 10.1099/0022-1317-69-12-2981
- Elde, N. C., Child, S. J., Eickbush, M. T., Kitzman, J. O., Rogers, K. S., Shendure, J., et al. (2012). Poxviruses deploy genomic accordeons to adapt rapidly against host antiviral defenses. *Cell* 150, 831–841. doi: 10.1016/j.cell.2012.05.049
- Funk, S. M., Verma, S. K., Larson, G., Prasad, K., Singh, L., Narayan, G., et al. (2007). The pygmy hog is a unique genus: 19th century taxonomists got it right first time round. *Mol. Phylogenet. Evol.* 45, 427–436. doi: 10.1016/j.ympev.2007.08.007
- Gabriel, C., Blome, S., Malogolovkin, A., Parilov, S., Kolbasov, D., Teifke, J. P., et al. (2011). Characterization of African swine fever virus Caucasus isolate in European wild boars. *Emerg. Infect. Dis.* 17, 2342–2345. doi: 10.3201/eid1712.110430
- Gallardo, C., Nurmoja, I., Soler, A., Delicado, V., Simón, A., Martin, E., et al. (2018). Evolution in Europe of African swine fever genotype II viruses from highly to moderately virulent. *Vet. Microbiol.* 219, 70–79. doi: 10.1016/j.vetmic.2018.04.001
- Gallardo, C., Soler, A., Nieto, R., Sanchez, M. A., Martins, C., Pelayo, V., et al. (2015). Experimental transmission of African Swine Fever (ASF) low virulent isolate NH/P68 by surviving pigs. *Transbound. Emerg. Dis.* 62, 612–622. doi: 10.1111/tbed.12431
- Golding, J. P., Goatley, L., Goodbourn, S., Dixon, L. K., Taylor, G., and Netherton, C. L. (2016). Sensitivity of African swine fever virus to type I interferon is linked to genes within multigene families 360 and 505. *Virology* 493, 154–161. doi: 10.1016/j.virol.2016.03.019
- Gómez del Moral, M., Ortuño, E., Fernández-Zapatero, P., Alonso, F., Alonso, C., Ezquerro, A., et al. (1999). African swine fever virus infection induces tumor necrosis factor alpha production: implications in pathogenesis. *J. Virol.* 73, 2173–2180.
- Graham, S. C., Bahar, M. W., Cooray, S., Chen, R. A., Whalen, D. M., Abrescia, N. G., et al. (2008). Vaccinia virus proteins A52 and B14 Share a Bcl-2-like fold but have evolved to inhibit NF-kappaB rather than apoptosis. *PLoS Pathog.* 4:e1000128. doi: 10.1371/journal.ppat.1000128
- Granberg, F., Torresi, C., Oggiano, A., Malmberg, M., Iscaro, C., De Mia, G. M., et al. (2016). Complete genome sequence of an african swine fever virus isolate from sardinia, Italy. *Genome Announc.* 4:e01220–16. doi: 10.1128/genomeA.01220-16
- Granja, A. G., Nogal, M. L., Hurtado, C., Del Aguila, C., Carrascosa, A. L., Salas, M. L., et al. (2006). The viral protein A238L inhibits TNF-alpha expression through a CBP/p300 transcriptional coactivators pathway. *J. Immunol.* 176, 451–462. doi: 10.4049/jimmunol.176.1.451
- Guinat, C., Gubbins, S., Vergne, T., Gonzales, J. L., Dixon, L., and Pfeiffer, D. U. (2016). Experimental pig-to-pig transmission dynamics for African swine fever virus, Georgia 2007/1 strain. *Epidemiol. Infect.* 144, 25–34. doi: 10.1017/S0950268815000862
- Hammond, R. A., and Detray, D. E. (1955). A recent case of African swine fever in Kenya, East Africa. *J. Am. Vet. Med. Assoc.* 126, 389–391.
- Hess, W. R., Endris, R. G., Haslett, T. M., Monahan, M. J., and McCoy, J. P. (1987). Potential arthropod vectors of African swine fever virus in North America and the Caribbean basin. *Vet. Parasitol.* 26, 145–155. doi: 10.1016/0304-4017(87)90084-7
- Hess, W. R., Endris, R. G., Lousa, A., and Caiado, J. M. (1989). Clearance of African swine fever virus from infected tick (Acari) colonies. *J. Med. Entomol.* 26, 314–317. doi: 10.1093/jmedent/26.4.314
- Hurtado, C., Granja, A. G., Bustos, M. J., Nogal, M. L., Gonzalez de Buitrago, G., de Yébenes, V. G., et al. (2004). The C-type lectin homologue gene (EP153R) of African swine fever virus inhibits apoptosis both in virus infection and in heterologous expression. *Virology* 326, 160–170. doi: 10.1016/j.virol.2004.05.019

- Jori, F., Vial, L., Penrith, M. L., Perez-Sanchez, R., Etter, E., Albina, E., et al. (2013). Review of the sylvatic cycle of African swine fever in sub-Saharan Africa and the Indian ocean. *Virus Res.* 173, 212–227. doi: 10.1016/j.virusres.2012.10.005
- Karalyan, Z., Zakaryan, H., Sargsyan, K., Voskanyan, H., Arzumanyan, H., Avagyan, H., et al. (2012). Interferon status and white blood cells during infection with African swine fever virus in vivo. *Vet. Immunol. Immunopathol.* 145, 551–555. doi: 10.1016/j.vetimm.2011.12.013
- Kipanyula, M. J., and Nong'ona, S. W. (2017). Variations in clinical presentation and anatomical distribution of gross lesions of African swine fever in domestic pigs in the southern highlands of Tanzania: a field experience. *Trop. Anim. Health Prod.* 49, 303–310. doi: 10.1007/s11250-016-1193-4
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kvansakul, M., Yang, H., Fairlie, W. D., Czabotar, P. E., Fischer, S. F., Perugini, M. A., et al. (2008). Vaccinia virus anti-apoptotic F1L is a novel Bcl-2-like domain-swapped dimer that binds a highly selective subset of BH3-containing death ligands. *Cell Death Differ.* 15, 1564–1571. doi: 10.1038/cdd.2008.83
- Leitao, A., Cartaxeiro, C., Coelho, R., Cruz, B., Parkhouse, R. M., Portugal, F., et al. (2001). The non-haemadsorbing African swine fever virus isolate ASFV/NH/P68 provides a model for defining the protective anti-virus immune response. *J. Gen. Virol.* 82(Pt 3), 513–523. doi: 10.1099/0022-1317-82-3-513
- Meijaard, E., Oliver, W. R. T., and Leus, K. (2017). *Sus cebifrons*. The IUCN Red List of Threatened Species 2017: e.T21175A44139575. doi: 10.1099/0022-1317-82-3-513
- Montgomery, R. E. (1921). On a form of swine fever occurring in British East Africa (Kenya colony). *J. Comp. Pathol. Ther.* 34, 159–191. doi: 10.1016/s0368-1742(21)80031-4
- Mujibi, F. D., Okoth, E., Cheruiyot, E. K., Onzere, C., Bishop, R. P., Fevre, E. M., et al. (2018). Genetic diversity, breed composition and admixture of Kenyan domestic pigs. *PLoS One* 13:e0190080. doi: 10.1371/journal.pone.0190080
- Munoz-Moreno, R., Cuesta-Geijo, M. A., Martinez-Romero, C., Barrado-Gil, L., Galindo, I., Garcia-Sastre, A., et al. (2016). Antiviral role of IFITM proteins in african swine fever virus infection. *PLoS One* 11:e0154366. doi: 10.1371/journal.pone.0154366
- Narayan, G., Deka, P., and Oliver, W. (2008). *Porcula salvania*. The IUCN Red List of Threatened Species 2008:e.T21172A9254675. doi: 10.1371/journal.pone.0154366
- Neidel, S., Maluquer de Motes, C., Mansur, D. S., Strnadova, P., Smith, G. L., and Graham, S. C. (2015). Vaccinia virus protein A49 is an unexpected member of the B-cell Lymphoma (Bcl)-2 protein family. *J. Biol. Chem.* 290, 5991–6002. doi: 10.1074/jbc.M114.624650
- Netherton, C. L., Simpson, J., Haller, O., Wileman, T. E., Takamatsu, H. H., Monaghan, P., et al. (2009). Inhibition of a large double-stranded DNA virus by MxA protein. *J. Virol.* 83, 2310–2320. doi: 10.1128/JVI.00781-08
- Nurmoja, I., Petrov, A., Breidenstein, C., Zani, L., Forth, J. H., Beer, M., et al. (2017). Biological characterization of African swine fever virus genotype II strains from north-eastern Estonia in European wild boar. *Transbound. Emerg. Dis.* 64, 2034–2041. doi: 10.1111/tbed.12614
- Oura, C. A. L., Powell, P. P., Anderson, E., and Parkhouse, R. M. E. (1998a). The pathogenesis of African swine fever virus in the resistant bushpig. *J. Gen. Virol.* 79, 1439–1443. doi: 10.1099/0022-1317-79-6-1439
- Oura, C. A. L., Powell, P. P., and Parkhouse, R. M. E. (1998b). African swine fever: a disease characterized by apoptosis. *J. Gen. Virol.* 79, 1427–1438. doi: 10.1099/0022-1317-79-6-1427
- Palgrave, C. J., Gilmour, L., Lowden, C. S., Lillico, S. G., Mellencamp, M. A., and Whitelaw, C. B. (2011). Species-specific variation in RELA underlies differences in NF-kappaB activity: a potential role in African swine fever pathogenesis. *J. Virol.* 85, 6008–6014. doi: 10.1128/JVI.00331-11
- Penrith, M. L., Thomson, G. R., Bastos, A. D., Phiri, O. C., Lubisi, B. A., Du Plessis, E. C., et al. (2004). An investigation into natural resistance to African swine fever in domestic pigs from an endemic area in southern Africa. *Rev. Sci. Tech.* 23, 965–977. doi: 10.20506/rst.23.3.1533
- Penrith, M. L., Vosloo, W., Jori, F., and Bastos, A. D. (2013). African swine fever virus eradication in Africa. *Virus Res.* 173, 228–246. doi: 10.1016/j.virusres.2012.10.011
- Petrov, A., Forth, J. H., Zani, L., Beer, M., and Blome, S. (2018). No evidence for long-term carrier status of pigs after African swine fever virus infection. *Transbound. Emerg. Dis.* 65, 1318–1328. doi: 10.1111/tbed.12881
- Plowright, W. (1977). "Vector transmission of African swine fever virus," in *Proceedings of the Agricultural Research Seminar of Hog Cholera/Classical Swine Fever and African Swine Fever EUR 5904* (Ixelles: European Commission of the European Communities), 575–587
- Plowright, W., Parker, J., and Peirce, M. A. (1969a). African swine fever virus in ticks (*ornithodoros moubata*, Murray) collected from animal burrows in Tanzania. *Nature* 221, 1071–1073. doi: 10.1038/2211071a0
- Plowright, W., Parker, J., and Pierce, M. A. (1969b). The epizootiology of African swine fever in Africa. *Vet. Rec.* 85, 668–674.
- Plowright, W., Perry, C. T., and Peirce, M. A. (1970). Transovarial infection with African swine fever virus in the argasid tick, *Ornithodoros moubata* porcinus, Walton. *Res. Vet. Sci.* 11, 582–584. doi: 10.1016/s0034-5288(18)34259-0
- Post, J., Weesendorp, E., Montoya, M., and Loeffen, W. L. (2017). Influence of age and dose of african swine fever virus infections on clinical outcome and blood parameters in pigs. *Viral Immunol.* 30, 58–69. doi: 10.1089/vim.2016.0121
- Quembo, C. J., Jori, F., Vosloo, W., and Heath, L. (2018). Genetic characterization of African swine fever virus isolates from soft ticks at the wildlife/domestic interface in Mozambique and identification of a novel genotype. *Transbound. Emerg. Dis.* 65, 420–431. doi: 10.1111/tbed.12700
- Reis, A. L., Abrams, C. C., Goatley, L. C., Netherton, C., Chapman, D. G., Sanchez-Cordon, P., et al. (2016). Deletion of African swine fever virus interferon inhibitors from the genome of a virulent isolate reduces virulence in domestic pigs and induces a protective response. *Vaccine* 34, 4698–4705. doi: 10.1016/j.vaccine.2016.08.011
- Reteno, D. G., Benamar, S., Khalil, J. B., Andreani, J., Armstrong, N., Klose, T., et al. (2015). Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *J. Virol.* 89, 6585–6594. doi: 10.1128/jvi.00115-15
- Rodríguez, J. M., Almazán, F., Viñuela, E., and Rodríguez, J. F. (1992). Genetic manipulation of African swine fever virus: construction of recombinant viruses expressing the beta-galactosidase gene. *Virology* 188, 67–76. doi: 10.1016/0042-6822(92)90735-8
- Rowlands, R. J., Duarte, M. M., Boinas, F., Hutchings, G., and Dixon, L. K. (2009). The CD2v protein enhances African swine fever virus replication in the tick vector, *Ornithodoros erraticus*. *Virology* 393, 319–328. doi: 10.1016/j.virol.2009.07.040
- Sánchez-Vizcaíno, J. M., Mur, L., Bastos, A. D. S., and Penrith, M. L. (2015). New insights into the role of ticks in African swine fever epidemiology. *Rev. Sci. Tech. Off. Int. Epiz.* 34:5030511.
- Sanna, G., Dei Giudici, S., Bacciu, D., Angioi, P. P., Giammarioli, M., De Mia, G. M., et al. (2017). Improved strategy for molecular characterization of African swine fever viruses from Sardinia, based on analysis of p30, CD2v and I73R/I329L variable regions. *Transbound. Emerg. Dis.* 64, 1280–1286. doi: 10.1111/tbed.12504
- Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., et al. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472, 481–485. doi: 10.1038/nature09907
- Souto, R., Mutowembwa, P., van Heerden, J., Fosgate, G. T., Heath, L., and Vosloo, W. (2016). Vaccine potential of two previously uncharacterized african swine fever virus isolates from southern africa and heterologous cross protection of an avirulent european isolate. *Transbound. Emerg. Dis.* 63, 224–231. doi: 10.1111/tbed.12250
- Tauscher, K., Pietschmann, J., Wernike, K., Teifke, J. P., Beer, M., and Blome, S. (2015). [On the situation of African swine fever and the biological characterization of recent virus isolates]. *Berl. Munch. Tierarztl. Wochenschr.* 128, 169–176.
- Thomas, L. F., Bishop, R. P., Onzere, C., McIntosh, M. T., Lemire, K. A., de Glanville, W. A., et al. (2016). Evidence for the presence of African swine fever virus in an endemic region of Western Kenya in the absence of any reported outbreak. *BMC Vet. Res.* 12:192. doi: 10.1186/s12917-016-0830-5
- Thompson, M. R., Kaminski, J. J., Kurt-Jones, E. A., and Fitzgerald, K. A. (2011). Pattern recognition receptors and the innate immune response to viral infection. *Viruses* 3, 920–940. doi: 10.3390/v3060920
- Thomson, G. R. (1985). The epidemiology of African swine fever: the role of free-living hosts in Africa. *Onderstepoort J. Vet. Res.* 52, 201–209.

- Thomson, G. R., Gainaru, M. D., and van Dellen, A. F. (1980). Experimental infection of warthog (*phacochoerus aethiopicus*) with African swine fever virus. *Onderstepoort J. Vet. Res.* 47, 19–22.
- Uttenthal, A., Braae, U. C., Ngowi, H. A., Rasmussen, T. B., Nielsen, J., and Johansen, M. V. (2013). ASFV in Tanzania: asymptomatic pigs harbor virus of molecular similarity to Georgia 2007. *Vet. Microbiol.* 165, 173–176. doi: 10.1016/j.vetmic.2013.01.003
- Whitworth, K. M., Rowland, R. R. R., Ewen, C. L., Tribble, B. R., Kerrigan, M. A., Cino-Ozuna, A. G., et al. (2016). Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus. *Nat. Biotech.* 34, 20–22. doi: 10.1038/nbt.3434
- Wilkinson, P. J. (1984). The persistence of ASF in Africa and the mediterranean. *Prevent. Vet. Med.* 2, 71–82. doi: 10.1016/0167-5877(84)90050-3
- Wilkinson, P. J., Pegram, R. G., Perry, B. D., Lemche, J., and Schels, H. F. (1988). The distribution of African swine fever virus isolated from *Ornithodoros moubata* in Zambia. *Epidemiol. Infect.* 101, 547–564. doi: 10.1017/s0950268800029423
- Xie, Z., Pang, D., Yuan, H., Jiao, H., Lu, C., Wang, K., et al. (2018). Genetically modified pigs are protected from classical swine fever virus. *PLoS Pathog.* 14:e1007193. doi: 10.1371/journal.ppat.1007193
- Zani, L., Forth, J. H., Forth, L., Nurmoja, I., Leidenberger, S., Henke, J., et al. (2018). Deletion at the 5'-end of Estonian ASFV strains associated with an attenuated phenotype. *Sci. Rep.* 8:6510. doi: 10.1038/s41598-018-24740-1
- Zhang, F., Moon, A., Childs, K., Goodbourn, S., and Dixon, L. K. (2010). The African swine fever virus DP71L protein recruits the protein phosphatase 1 catalytic subunit to dephosphorylate eIF2alpha and inhibits CHOP induction but is dispensable for these activities during virus infection. *J. Virol.* 84, 10681–10689. doi: 10.1128/JVI.01027-10

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Netherton, Connell, Benfield and Dixon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Quantitative Trait Loci Mapping for Lameness Associated Phenotypes in Holstein–Friesian Dairy Cattle

Enrique Sánchez-Molano^{1†}, Veyssel Bay^{2,3†}, Robert F. Smith⁴, Georgios Oikonomou^{2,4*} and Georgios Banos^{1,5*‡}

¹ The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush, Edinburgh, United Kingdom, ² Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom, ³ Bandirma Sheep Research Institute, The Ministry of Agriculture and Forestry, Balıkesir, Turkey, ⁴ Institute of Veterinary Science, University of Liverpool, Leahurst Campus, Liverpool, United Kingdom, ⁵ The Roslin Institute Building, Scotland's Rural College, Easter Bush, Edinburgh, United Kingdom

OPEN ACCESS

Edited by:

John Anthony Hammond,
Pirbright Institute,
United Kingdom

Reviewed by:

Hermann H. Swalve,
Martin Luther University of
Halle-Wittenberg, Germany
Filippo Biscarini,
Italian National Research Council
(CNR), Italy

*Correspondence:

Georgios Banos
Georgios.Banos@roslin.ed.ac.uk

[†]These authors shared first
authorship

[‡]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 10 September 2018

Accepted: 05 September 2019

Published: 04 October 2019

Citation:

Sánchez-Molano E, Bay V, Smith RF,
Oikonomou G and Banos G (2019)
Quantitative Trait Loci Mapping for
Lameness Associated Phenotypes in
Holstein–Friesian Dairy Cattle.
Front. Genet. 10:926.
doi: 10.3389/fgene.2019.00926

Lameness represents a significant challenge for the dairy cattle industry, resulting in economic losses and reduced animal health and welfare. The existence of underlying genomic variation for lameness associated traits has the potential to improve selection strategies by using genomic markers. Therefore, the aim of this study was to identify genomic regions and potential candidate genes associated with lameness traits. Lameness related lesions and digital cushion thickness were studied using records collected by our research team, farm records, and a combination of both. Genome-wide analyses were performed to identify significant genomic effects, and a combination of single SNP association analysis and regional heritability mapping was used to identify associated genomic regions. Significant genomic effects were identified for several lameness related traits: Two genomic regions were identified on chromosome 3 associated with digital dermatitis and interdigital hyperplasia, one genomic region on chromosome 23 associated with interdigital hyperplasia, and one genomic region on chromosome 2 associated with sole haemorrhage. Candidate genes in those regions are mainly related to immune response and fibroblast proliferation. Quantitative trait loci (QTL) identified in this study could enlighten the understanding of lameness pathogenesis, providing an opportunity to improve health and welfare in dairy cattle with the addition of these regions into selection programs.

Keywords: Lameness, GWAS, welfare, regional heritability mapping, QTL

INTRODUCTION

Lameness is a complex trait defined as an abnormal stance or gait of the animal that results from disorders of the locomotor system. In dairy cattle, lameness is one of the most important health conditions together with impaired fertility and mastitis (Green et al., 2002; Buitenhuis et al., 2007; Cha et al., 2010), and causes important economic losses and reduced animal health and welfare (Huxley, 2013).

Many lameness cases are associated to various infectious and non-infectious diseases (Green et al., 2002; Van Der Waaij et al., 2005; Bicalho and Oikonomou, 2013), resulting in painful foot lesions such as sole ulcers, white line lesions, sole haemorrhages, interdigital hyperplasia, and others. Previous studies have shown that both animal genetic and management factors contribute to the

development of these diseases (Olmos et al., 2009; Van Der Linde et al., 2010; Swalve et al., 2014). The existence of genetic variation underlying lameness-associated traits has been previously demonstrated using pedigree data analyses, with heritability estimates ranging from 0.06 to 0.52 (Boettcher et al., 1998; Zwald et al., 2004; Koenig et al., 2005; Laursen et al., 2009; Schopke et al., 2015). Furthermore, susceptibility to certain non-infectious foot lesions is also associated with morphological hoof traits such as the thickness of the digital cushion, a complex, force dissipating, subcutaneous tissue located under the distal phalanx (Bicalho et al., 2009).

While reducing the incidence of lameness is one of the main objectives for the dairy cattle industry, current evaluations are based on observational scores such as claw health status data, lameness and mobility scores, conformational traits, and data collected by automated sensors (Heringstad et al., 2018). All this information is obtained once the animal has started to show symptoms of lameness, thus not being available early in life and also showing relatively low heritabilities (Buitenhuis et al., 2007). Therefore, the identification of genomic regions and genes associated to lameness and lameness associated traits could strongly improve selection strategies by providing genomic information to make early breeding decisions and potentially informing more accurate genomic-based selection programmes.

Few studies have addressed traits associated with lameness using a genomic approach. The largest study (Malchiodi et al., 2018) grouped lameness-associated lesions into two categories, infectious and non-infectious, and used single nucleotide polymorphism (SNP) data to identify several genomic regions with candidate genes linked to immune system, morphogenesis, and cell proliferation. Other studies used microsatellite data (Buitenhuis et al., 2007) to identify lameness-associated regions and SNP data (Scholey et al., 2012; Swalve et al., 2014) to identify regions associated to digital dermatitis and sole hemorrhage. The relatively high number of identified regions together with the complex aetiology of lameness seems to support a potential polygenic architecture with many genes influencing the different biological factors involved. Therefore, it is necessary to study the different types of lesions separately in order to identify particular and common genomic regions that contribute to the main condition phenotype.

The objective of the present study is to perform genome-wide analyses to identify regions and candidate genes, and understand the genetic basis of a wide range of lameness-related traits. This knowledge may inform genetic improvement schemes aiming to reduce prevalence of dairy cattle lameness. Digital cushion thickness (DCT) measurements are studied here for the first time from a genomic perspective.

MATERIALS AND METHODS

Animals and Phenotypes

Ethical approval for the study was granted by the University of Liverpool Research Ethics Committee. ASPA regulated procedures were conducted under a Home Office License (Reference Number: PPL 70/8330).

The study included a total of 554 Holstein–Friesian cows in lactation 0–8 from three different farms. Data sources were from a combination of previous research projects and routinely kept farm records. The recorded lameness-causing foot lesions were digital dermatitis, sole ulcer, white line disease, sole haemorrhage, and interdigital hyperplasia. Cases were defined following the ICAR Claw Health Atlas definitions (Egger-Danner et al., 2015).

Farm phenotypic records for presence (1) or absence (0) of these lesions were extracted from the farm database for all these animals (single record per animal) from May 2006 to October 2017 using TotalVet software (Sum-IT). In addition, 475 cows were individually monitored for the same lesions by a research team led by an experienced veterinarian during three separate time intervals between December 2014 and October 2017. DCT measurements were taken using ultrasonography between December 2014 and January 2016 (1st research interval), and between October 2016 and August 2017 (2nd research interval). A number of cows were also followed during the period between August 2017 and October 2017, but no DCT measurements were obtained (3rd research interval). Lameness lesions were recorded by the veterinarian for 88 animals between December 2014 and January 2016, and for 337 animals between October 2016 and October 2017; 50 cows had records from both these two different time intervals. DCT measurements and recording of lameness-causing lesions were performed at three time points around animals' calving: 3–4 weeks before the expected calving, the first week after calving, and approximately 8 weeks after calving. Seventy-nine animals only had lameness lesion records obtained from the farms' records. Research and farm records were analysed both separately and combined. For the latter, animals were considered as affected when at least one of the available records (research or farm) indicated presence of the lesion. All these data are summarised in **Table 1**.

Cows were restrained in a foot trimming crush for the measurement of DCT and the recording of lameness-causing foot lesions. Measurement of DCT was performed using an Easi-Scan ultrasound machine (sonographic B-mode, BCFTM Technology, UK) equipped with a linear probe 5–8 MHz. All

TABLE 1 | Number of cows enrolled *per* farm and source of data collection.

| Farm | n | Source of data collection | n |
|------|-----|---------------------------|-----|
| 1 | 307 | 0 | 79 |
| | | 1 | 88 |
| | | 2 | 81 |
| | | 3 | 9 |
| | | 4 | 45 |
| 2 | 135 | 5 | 5 |
| | | 2 | 74 |
| | | 3 | 61 |
| 3 | 112 | 2 | 112 |

Source of data collection: 0 = farm records only, 1 = research data collected during 1st research interval (December 2014 and January 2016), 2 = research data collected during 2nd research interval (October 2016 and August 2017), 3 = research data collected during 3rd research interval (August 2017 and October 2017), 4 = research data collected during 1st and 2nd research intervals, 5 = research data collected during 1st and 3rd research intervals.

measurements of DCT were undertaken at the midline, on the lateral claw of the hind left foot. To measure the DCT, the foot was cleaned and loose horn was removed with a hoof knife. Sole contact with the transducer was made using ultrasound gel (Ultrasound Gel, Henry Schein) and a gel standoff (Flexi gel standoff, BCFTM Technology, UK). After freezing the image on the ultrasound monitor (Easi-Scan Ultrasound Remote Display, BCFTM Technology, UK), measurements were taken to the nearest millimetre. The DCT was measured just cranial to the tuberculum flexorum of the pedal bone at the typical sole ulcer site. The distance from the inner margin of the sole (identified as a thin echogenic line) to the distal edge of the pedal bone (identified as a thick echogenic line) was assessed.

DNA Sampling, Extraction, and Genotyping

Blood samples were collected from the tail vein of each cow in vacutainer tubes containing EDTA. Genomic DNA was extracted from buffy coat samples using the QIAamp DNA Blood MiniKit from Qiagen. Extracted DNA samples were quantified using a NanoDrop and stored at -20°C . Initially, 266 cows were genotyped using the Affymetrix Axiome bovine 54K SNP array. The Illumina BovineSNP50 bead chip containing 53,714 SNP was used to genotype the rest of the animals. Genotype data obtained from the Affymetrix array were converted to the Illumina chip format by selecting common SNPs using concordant strand assignment and identified allelic calls before further analyses were conducted.

Sample and Genotype Quality Control

Quality control was performed using PLINK (Purcell et al., 2007) in order to assess both sample and marker quality. A minimum genotype call rate of 95% was applied, removing SNPs with low genotyping quality. Further quality control on the markers removed those with low minor allele frequency ($\text{MAF} < 0.01$) and showing strong deviations from Hardy-Weinberg equilibrium (threshold of $1.45\text{E-}6$ calculated genome-wide by applying a Bonferroni correction to obtain a nominal P -value of 0.05). The final genotype call rate after genotype quality control was 98.7%. Additional quality control of samples was performed by removing individuals with poor genotype quality (sample call rate lower than 95%). All these quality control procedures resulted in a final dataset of 549 animals genotyped for 34,658 SNPs with positions assigned according to the UMD 3.1 assembly.

Population Structure

Principal component analyses of the genotyped animals were performed using GEMMA (Zhou and Stephens, 2012), replacing missing genotypes with the average genotype. Visual exploration showed a relatively light population structure not explained by any of the descriptive factors (e.g., farm, parity number, lactation, etc.). Therefore, the genomic relationship matrix (GRM) among animals was fitted in all ensuing statistical models of analysis as a random polygenic effect to account for any potential inflation effects caused by population structure.

Estimation of Variance Components

Estimates of the variance components for each individual trait were obtained by fitting the following model using REACTA (Cebamanos et al., 2014):

$$\mathbf{y} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{y} represents the vector of phenotypes, \mathbf{W} is an incidence matrix, $\boldsymbol{\alpha}$ is the vector of associated fixed effects, \mathbf{Z} is the design matrix for the vector \mathbf{u} of random polygenic effects [distributed as a multivariate normal distribution $\text{MVN}(0, \mathbf{V}_g \mathbf{G})$ with \mathbf{G} being the GRM and \mathbf{V}_g the genetic variance of the trait], and $\boldsymbol{\varepsilon}$ represents the vector of residual errors [distributed as $\text{MVN}(0, \mathbf{V}_e \mathbf{I})$ with \mathbf{I} being the identity matrix and \mathbf{V}_e the residual variance]. The significance of the genomic (polygenic) effect ($P = 0.05$) was assessed using the likelihood ratio test statistic to compare a model that fits the effect against the base model that excludes it.

Fixed effects used in model (1) were tested previously using Wald tests in ASReml 4 (Gilmour et al., 2009), fitting a logit model for disease traits and a linear model for DCT records, and following a backward elimination approach. After performing analyses for all traits, concordant models were chosen incorporating as fixed effects: i) for the disease research records: farm (3 levels), parity number at recording (three levels, 1, 2, and ≥ 3), and research interval (five levels, grouped as 1 = interval 1, 2 = interval 2, 3 = interval 3, 4 = intervals 1 and 2, 5 = intervals 1 and 3); ii) for the farm and combined disease records: farm (as before), lactation number at the end of study (four levels 0, 1, 2, and >3), and interval (as before); and iii) for the DCT records: farm (as before), parity number (as before), and assessor (six levels).

Genome-Wide Association Analysis

Individual SNP association analyses were performed in those traits with a significant genomic effect from model (1) using GEMMA (Zhou and Stephens, 2012). The linear mixed model was:

$$\mathbf{y} = \mathbf{W}\boldsymbol{\alpha} + \mathbf{x}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2)$$

where \mathbf{x} represents the vector of genotypes (coded as 0/1/2) and $\boldsymbol{\beta}$ is the regression coefficient of the phenotype on the genotypes; all other effects are as described in model (1). The statistical significance of the regression coefficient was assessed using a Wald test. When determining the significant thresholds, a Bonferroni correction was performed for multiple testing due to the number of markers, but not for multiple traits. This resulted in a genome-wide significant threshold ($P = 0.05$) defined at $P = 1.44\text{E-}6$ [$-\log_{10}(P) = 5.84$] and a suggestive threshold (one false positive *per* genome scan, $P = 1$) defined at $P = 2.89\text{E-}5$ [$-\log_{10}(P) = 4.54$].

Despite including the polygenic effect in the model, genotyping errors or other artefacts such as cryptic population structure may inflate test statistics. Therefore, to account for any potential remaining inflation, the ratio of the median of the empirically

observed distribution of the test statistic to the expected median (inflation factor λ) was used for correction, following the method described by Amin et al. (2007), which assumes that the inflation is constant across the genome.

Regional Heritability Mapping

Under the regional heritability mapping approach, the genome was divided into non-overlapping windows of 20 consecutive SNPs. The following model was used in REACTA (Cebamanos et al., 2014):

$$\mathbf{y} = \mathbf{W}\alpha + \mathbf{X}\mathbf{u}_{(i)} + \mathbf{Z}\mathbf{u}_{(-i)} + \varepsilon \quad (3)$$

where \mathbf{X} and \mathbf{Z} are the corresponding design matrices for the effects $\mathbf{u}_{(i)}$ of the corresponding region i {distributed as $\text{MVN}[0, V_{g(i)}\mathbf{G}_{(i)}]$, with $V_{g(i)}$ and $\mathbf{G}_{(i)}$ being the genomic variance and the GRM corresponding to the SNPs in the i^{th} region, respectively} and $\mathbf{u}_{(-i)}$ of the genome (polygenic effect) excluding the region i {distributed as $\text{MVN}[0, V_{g(-i)}\mathbf{G}_{(-i)}]$ with $V_{g(-i)}$ and $\mathbf{G}_{(-i)}$ being the genetic variance and the GRM corresponding to all SNPs other than those on the region i , respectively}.

The significance of the region effect was assessed using the likelihood ratio test statistic. A total of 1,733 regions were analysed, leading to a genome-wide significant threshold ($P = 0.05$) defined at $P = 2.89\text{E-}5$ with Bonferroni correction for multiple regions [$-\log_{10}(P) = 4.54$] and a suggestive threshold (one false positive *per* genome scan) defined at $P = 5.77\text{E-}4$ [$-\log_{10}(P) = 3.24$]. As with the genome-wide association analyses, a correction by the inflation factor λ was applied to account for any remaining inflation after fitting the polygenic effect in the model.

RESULTS

Descriptive Statistics

The number of cows enrolled *per* farm and *per* period of data collection is presented in Table 1. Incidences of each lameness-inducing foot lesion *per* farm are summarised in Table 2. Number of cows *per* farm with no, one, two, or three or more of the recorded lesions are presented in Table 3.

Population Structure

Figure 1A shows the eigenvalues corresponding to the principal component analysis performed on the GRM of the genotyped animals. The first seven principal components accounted for about 10% of the total variance, with the first three components

TABLE 3 | Number of cows *per* farm with zero, one, two, and three or more recorded lesions (digital dermatitis, interdigital hyperplasia, sole haemorrhage, sole ulcer, and white line disease)

| Farm | Number of lesions <i>per</i> cow | n |
|------|----------------------------------|----|
| 1 | 0 | 86 |
| | 1 | 82 |
| | 2 | 69 |
| | 3+ | 70 |
| 2 | 0 | 21 |
| | 1 | 55 |
| | 2 | 31 |
| | 3+ | 28 |
| 3 | 0 | 44 |
| | 1 | 48 |
| | 2 | 14 |
| | 3+ | 6 |

explaining 2.19%, 1.83%, and 1.50%, respectively. Figure 1B shows a light population structure mainly due to the first principal component. No population attributes were available that could explain this. Therefore, a polygenic effect based on GRM was fitted to account for this light population structure.

Full Genomic Variance Analysis

Table 4 shows the variance component estimates in the observed scale for those traits with a significant genomic effect ($P < 0.05$) based on the likelihood ratio test. Heritability estimates for the disease traits range from 0.129 (white line disease in combined records) to 0.516 (interdigital hyperplasia in research records), corresponding to genomic variances in the range of 0.008–0.067. Heritability for DCT at calving was moderate (0.228), corresponding to a genetic variance of 0.549.

Genome-Wide Association Analysis

Two genome-wide significant SNPs (for interdigital hyperplasia and digital dermatitis in farm records) and 19 genome-wide suggestive SNPs were detected in the genome-wide association analyses (Table 5). After performing the correction by the inflation factor, all λ estimates ranged from 1.002 to 1.033, thus implying the absence of any significant inflation in the test estimates.

MAFs ranged from 0.020 to 0.479, and most substitution effects were positive (with the exception of white line disease in the combined records), thus implying a positive effect of the minor allele against the disease. However, due to a possible Beavis effect (Xu, 2003), the provided effect sizes may be slightly overestimated.

DCT at calving and sole ulcer in the combined records did not provide any significant or suggestive SNP despite showing a significant genomic effect in the previous analyses.

Regional Heritability Mapping and Concordant Regions

A significant region for interdigital hyperplasia was detected based on farm records, and 10 suggestive regions were detected for other

TABLE 2 | Incidence of digital dermatitis (DD), interdigital hyperplasia (IH), sole haemorrhage (SH), sole ulcer (SU), and white line disease (WLD) *per* farm.

| Farm | SU | SH | WLD | DD | IH |
|------|------|------|------|------|------|
| 1 | 0.26 | 0.28 | 0.32 | 0.40 | 0.23 |
| 2 | 0.27 | 0.42 | 0.24 | 0.57 | 0.07 |
| 3 | 0.11 | 0.09 | 0.07 | 0.46 | 0.12 |

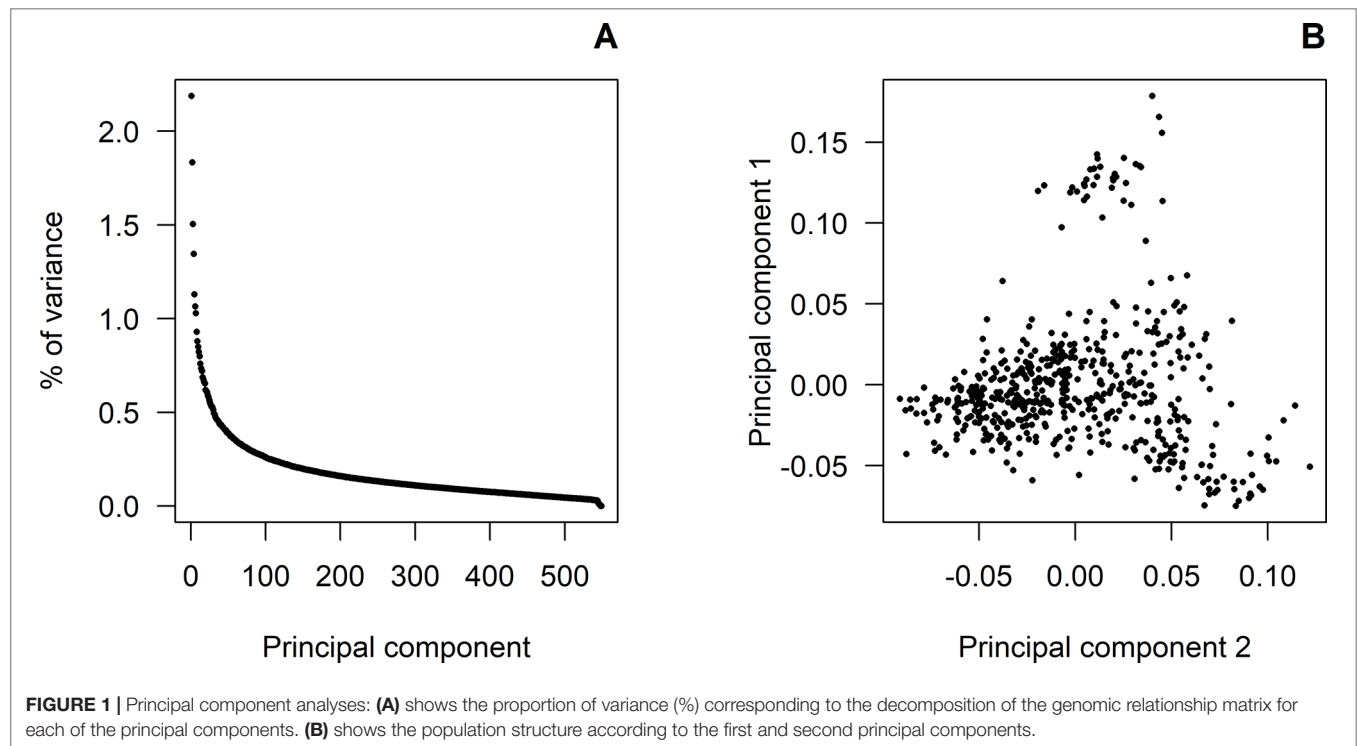


TABLE 4 | Estimates of heritability and variance components for traits with a significant ($P < 0.05$) genomic effect.

| | Trait | h^2 | V_g | V_e | P | N |
|-------------------------|------------------|-------------------|-------------------|-------------------|-----------|-----|
| DCT records | DCT_fresh | 0.228 ± 0.119 | 0.549 ± 0.298 | 1.854 ± 0.290 | $6.69E-3$ | 360 |
| | DD | 0.185 ± 0.093 | 0.043 ± 0.022 | 0.189 ± 0.023 | $5.55E-3$ | 469 |
| Research records | IH | 0.516 ± 0.105 | 0.067 ± 0.016 | 0.063 ± 0.012 | $1.05E-8$ | 469 |
| | SH | 0.204 ± 0.099 | 0.035 ± 0.017 | 0.135 ± 0.017 | $1.25E-2$ | 469 |
| Farm records | DD | 0.196 ± 0.088 | 0.023 ± 0.011 | 0.094 ± 0.011 | $5.89E-3$ | 549 |
| | SU | 0.290 ± 0.091 | 0.037 ± 0.013 | 0.092 ± 0.011 | $6.29E-5$ | 549 |
| | IH | 0.131 ± 0.076 | 0.008 ± 0.005 | 0.054 ± 0.005 | $1.85E-2$ | 549 |
| | DD | 0.201 ± 0.080 | 0.046 ± 0.019 | 0.183 ± 0.019 | $2.56E-4$ | 549 |
| | SU | 0.350 ± 0.096 | 0.057 ± 0.017 | 0.106 ± 0.015 | $8.24E-6$ | 549 |
| Combined records | WLD | 0.129 ± 0.081 | 0.020 ± 0.013 | 0.135 ± 0.014 | $2.83E-2$ | 549 |
| | IH | 0.368 ± 0.088 | 0.047 ± 0.013 | 0.081 ± 0.011 | $4.42E-8$ | 549 |
| | SH | 0.143 ± 0.083 | 0.024 ± 0.014 | 0.142 ± 0.015 | $3.19E-2$ | 549 |

Genomic heritabilities (h^2) and genomic (V_g) and residual variances (V_e) estimated together with their standard errors. P -values (P) for the significance of the genomic effect and the number of total records (N). Digital cushion thickness at calving (DCT_fresh), digital dermatitis (DD), interdigital hyperplasia (IH), sole haemorrhage (SH), sole ulcer (SU), and white line disease (WLD).

traits using the regional heritability mapping approach (Table 6). The significant region detected on chromosome 3 for digital dermatitis using research records was also detected as suggestive for interdigital hyperplasia using the combined records.

Four of the detected regions were concordant with suggestive/significant SNPs detected in the genome-wide association analyses (Table 5 and Supplementary Figure 1), two of them detected from the farm records and another two from the combined farm and research records. In the farm records, the concordant regions detected both by genomic analyses explained 32.90% and 43.90% of the total genomic variance of digital dermatitis and interdigital hyperplasia, respectively. In the combined records, the concordant regions explained 10.53% and

11.29% of the total genomic variance of interdigital hyperplasia and sole haemorrhage, respectively. Again, caution must be exercised while assessing these findings because of possible overestimation due to a potential Beavis effect (Xu, 2003).

DISCUSSION

In the present study, two genome-wide association approaches were used to identify quantitative trait loci (QTL) affecting lameness related traits. Comparison of the results provided by individual SNP genome-wide association analyses and regional heritability mapping was performed to strengthen the evidence of

TABLE 5 | Significant SNP from the genome-wide association analyses.

| | Trait | BTA | Position (BP) | Beta coef. | MAF | P |
|-------------------------|------------|-----|---------------|----------------|-------|---------|
| Research records | DD | 3 | 70931186 | 0.380 ± 0.086 | 0.038 | 1.23E-5 |
| | | 20 | 30216498 | 0.166 ± 0.037 | 0.276 | 7.74E-6 |
| | IH | 11 | 99952182 | 0.180 ± 0.042 | 0.098 | 2.54E-5 |
| | SH | 2 | 4958110 | 0.170 ± 0.038 | 0.171 | 1.00E-5 |
| Farm records | DD | 3 | 70931186 | 0.275 ± 0.056 | 0.038 | 1.32E-6 |
| | | 7 | 28258117 | 0.322 ± 0.074 | 0.020 | 1.97E-5 |
| | | 19 | 10140328 | 0.264 ± 0.062 | 0.027 | 2.91E-5 |
| | | 24 | 37354445 | 0.137 ± 0.031 | 0.147 | 1.32E-5 |
| | SU | 12 | 12612422 | 0.185 ± 0.043 | 0.084 | 2.82E-5 |
| | IH | 23 | 44153826 | 0.266 ± 0.048 | 0.027 | 4.83E-8 |
| Combined records | DD | 3 | 70931186 | 0.383 ± 0.080 | 0.038 | 2.17E-6 |
| | | 3 | 90367814 | 0.224 ± 0.047 | 0.123 | 2.22E-6 |
| | | 3 | 90523019 | 0.239 ± 0.053 | 0.096 | 7.42E-6 |
| | | 20 | 30216498 | 0.165 ± 0.035 | 0.270 | 2.30E-6 |
| | WLD | 5 | 94496854 | -0.106 ± 0.025 | 0.479 | 2.31E-5 |
| | | 7 | 75190535 | -0.122 ± 0.027 | 0.339 | 4.99E-6 |
| | | 14 | 5883219 | -0.148 ± 0.034 | 0.152 | 1.60E-5 |
| | IH | 2 | 23628756 | 0.157 ± 0.037 | 0.106 | 2.07E-5 |
| | | 3 | 23764339 | 0.160 ± 0.037 | 0.119 | 1.61E-5 |
| | SH | 2 | 4958110 | 0.152 ± 0.035 | 0.165 | 1.93E-5 |
| | | 21 | 46018333 | 0.262 ± 0.057 | 0.052 | 4.94E-6 |

Chromosome (BTA) and base pair position follow the UMD 3.1 assembly. Beta coefficient (minor allele substitution effect) and standard error, minor allele frequency (MAF), and P-value (P) for the beta coefficient. Digital dermatitis (DD), interdigital hyperplasia (IH), sole haemorrhage (SH), sole ulcer (SU), and white line disease (WLD).

TABLE 6 | Significant genomic regions from the regional heritability mapping analyses.

| | Trait | BTA | Position (BP) | P |
|-------------------------|------------|-----|-------------------|---------|
| Research records | DD | 22 | 34216267–36047120 | 4.14E-4 |
| | | 22 | 31087678–32538832 | 4.36E-4 |
| Farm records | DD | 3 | 70077512–71882823 | 3.40E-4 |
| | | 25 | 34887253–35853810 | 2.61E-4 |
| | SU | 25 | 3126438–4354023 | 8.03E-5 |
| | IH | 23 | 43151282–44458259 | 2.28E-6 |
| Combined records | SU | 25 | 3126438–4354023 | 1.86E-4 |
| | WLD | 14 | 6850767–7718808 | 5.43E-4 |
| | IH | 3 | 22069239–23764339 | 1.35E-4 |
| | | 3 | 70077512–71882823 | 5.94E-5 |
| | SH | 2 | 4587203–5640288 | 3.35E-4 |

Chromosome (BTA) and base pair position follow the UMD 3.1 assembly. P-value (P) for the region effect. Digital dermatitis (DD), interdigital hyperplasia (IH), sole haemorrhage (SH), sole ulcer (SU), and white line disease (WLD).

the identified regions. QTLs were detected for digital dermatitis, interdigital hyperplasia, and sole haemorrhage.

Three sources of data were used in this study in order to identify genomic regions for lameness related diseases. Although using only research-confirmed records is expected to provide more accurate phenotypes than using farm records, the number of observations available was smaller, thus reducing the power to detect significant genomic effects. Similarly, using farm records provided a larger number of records spanning animals' whole lifetime; these records however are potentially less accurate, thus leading to a low detection power and increase the chances of introducing misclassification bias. Impact of phenotypic errors in genomic analyses has been

discussed before both in the context of human (Buyske et al., 2009) and animal data (Biffani et al., 2017). Furthermore, it has been shown previously that farm records could seriously under-record certain lesions (Heringstad et al., 2018) and this has also been the case with our dataset. The most powerful dataset available in the present study was the combination of farm and research records, which provided a larger number of records than research alone but more accurate compared to farm, thus leading to significant estimates of genomic effects for more traits.

Heritability estimates for some of these traits have been previously estimated using pedigree data (Koenig et al., 2005; Van Der Waaij et al., 2005; Gernand et al., 2012; Oberbauer et al., 2013; Van Der Spek et al., 2013; Malchiodi et al., 2017). Such estimates range from 0.07 to 0.4 for digital dermatitis, 0.10 to 0.39 for interdigital hyperplasia, and 0.04 to 0.17 for sole haemorrhage. Our heritability estimates were generally in concordance within these ranges, particularly considering the estimates obtained using the combined records. It has to be recognised that heritability estimates are presented in the observed scale (0–1) and, therefore, population parameters are dependent on the disease prevalence (Lee et al., 2011). However, it is expected that estimates will not vary widely when transformed to the liability scale (normal distribution), where the mean and variance are independent from the prevalence of the disease. In the case of digital dermatitis, the heritability observed for the combined records was 0.20, resulting in estimates between 0.21 and 0.30 on the liability scale when assuming a disease prevalence from 10% to 30% (Holzhauer et al., 2006). Similarly, the heritability observed for interdigital hyperplasia was 0.37, resulting in a heritability of 0.39 on the liability scale when assuming a prevalence of 1.3% (Solano et al., 2016).

Although genome-wide association and regional heritability mapping analyses revealed several QTLs independently, four QTLs were commonly reported by both approaches (**Tables 2 and 3 and Supplementary Figure 1**). On chromosome 3, a suggestive region was associated with digital dermatitis in farm records, explaining 32.90% of the total genomic variance and being also suggestive of interdigital hyperplasia in the combined records. Two potential gene candidates are contained within this region: i) *FPGT* (fucose-1-phosphate guanylyl transferase) part of the L-fucose pathway, a key sugar in complex carbohydrates involved in cell-to-cell recognition, inflammation, and immune processes (Becker and Lowe, 2003); and ii) *TNNI3K* (serine/threonine-protein kinase TNNI3K), also associated with inflammation mechanisms (Wiltshire et al., 2011). Based on this function, we surmise that a candidate gene for lameness resistance may be found within this QTL.

On chromosome 23, a significant region was associated with interdigital hyperplasia in farm records, explaining 43.90% of the total genomic variance. Interdigital hyperplasia, also known as interdigital fibroma (Atkinson, 2013), results in a thickening of interdigital connective tissue causing fibroid tumours. Thus, a potential candidate gene found within this region is *EDN1* (endothelin-1), a vasoconstrictor associated with several cardiovascular diseases and inflammatory and fibrotic processes (Matsushima et al., 2004), acting as fibroblast mitogen in systemic sclerosis (Vancheeswaran et al., 1994), pulmonary fibrosis (Hoher et al., 2000), and hepatic fibrosis (Rockey and Chung, 1996).

On chromosome 3, another suggestive region was associated with interdigital hyperplasia using combined records, explaining 10.53% of the total genomic variance. This region includes several potential candidate genes, particularly *PHGDH* (D-3-phosphoglycerate dehydrogenase), an oxidoreductase that has been associated previously with pulmonary fibrosis (Hamanaka et al., 2017).

On chromosome 2, a suggestive region was associated with sole haemorrhage from the combined records, explaining 11.29% of the total genomic variance. With sole haemorrhage being related to impaired vascular system and cellular inflammatory reactions (Ossent and Lischer, 1998), a potential candidate gene within this region is *GPR17* (uracil nucleotide/cysteinyl leukotriene receptor). This gene is as a sensor molecule involved in traumatic, vascular, and inflammatory pathologies in the central nervous system (Boda et al., 2011), and is also related to vascular permeability and inflammatory processes as a regulator of the cysteinyl leukotriene 1 receptor response (Maekawa et al., 2009).

Most candidate genes within the detected genomic regions are related either to inflammatory processes or fibroblast proliferation, as expected due to the nature of the analysed traits. However, these are not independent processes, but linked networks where fibroblasts present complex biosynthetic pathways, playing a role in pathogenesis and mediating inflammatory processes through their proliferation (Smith, 2005). Thus, the analysed traits are expected to present a complex genomic architecture with several genes and pathways involved in their phenotypic expression. This is concordant

with the possible overestimation of the SNP and regional effects due to the Beavis effect (Xu, 2003) as well as with the lack of consistency across QTLs detected in several studies (Buitenhuis et al., 2007; Scholey et al., 2012; Swalve et al., 2014; Malchiodi et al., 2018). Therefore, it is expected that increasing the sample size with additional accurate records will increase the accuracy of the estimates of these effects.

DCT is a novel trait analysed using genomic data for the first time in the present study. Previous studies have shown an association between lameness-related diseases such as sole ulcer and white line disease with a thinner digital cushion, indicating also a potential change in the tissue composition of the cushion (Bicalho et al., 2009). In our study, given the relatively small number of samples available, no significant or suggestive markers were identified. However, a significant genomic effect was detected for DCT at calving, providing a moderate genomic heritability of 0.23 ± 0.12 . When compared with the heritability of 0.33 obtained in a previous pedigree-based study (Oikonomou et al., 2014), our estimate was smaller but within the standard error boundaries.

As with the lameness-associated lesions, the genomic architecture of DCT traits is expected to be polygenic, being particularly related with body fatty acid and lipid metabolism. Further studies with an increased sample size will refine the heritability estimates and provide some potential candidate genes associated with this structure.

CONCLUSION

Four genomic regions were identified for digital dermatitis, interdigital hyperplasia, and sole haemorrhage, harbouring genes involved in inflammatory and fibroblastic processes. These traits are moderately heritable and potentially associated with a polygenic architecture. Therefore, the identification of associated regions may be useful to inform genomic selection programmes against lameness and to increase our knowledge of the underlying pathology.

In addition, this is the first study to address DCT from a genomic perspective, showing a moderate genomic heritability for this structure during the period of calving. The genomic architecture of this trait warrants further research attention.

DATA AVAILABILITY STATEMENT

The genotype data has been uploaded to a public repository hosted by the University of Edinburgh. Genotypes are therefore publicly available and can be obtained from: Edinburgh DataShare (University of Edinburgh), <https://datashare.is.ed.ac.uk/handle/10283/3409>.

ETHICS STATEMENT

Ethical approval for the study was granted by the University of Liverpool Research Ethics Committee. ASPA regulated

procedures were conducted under a Home Office License (Reference Number: PPL 70/8330).

AUTHOR CONTRIBUTIONS

ES-M analyzed the data and co-wrote the first draft of the manuscript. VB collected the data, performed laboratory work, analyzed the data, and co-wrote the first draft of the manuscript. RS provided funding and critically evaluated the manuscript. GO and GB secured funding, designed and supervised the study, and critically evaluated the manuscript.

FUNDING

This study was funded by the Academy of Medical Sciences. VB acknowledges support from The Turkish Ministry of Education, and Ministry of Food, Agriculture, and Livestock. GO gratefully acknowledges support from the Wellcome Trust. GB and ES-M gratefully acknowledge funding from The Roslin Institute Strategic Programme (ISP) grants, and GB also the Rural and Environment Science and Analytical Services Division of the Scottish Government.

REFERENCES

- Amin, N., Van Duijn, C. M., and Aulchenko, Y. S. (2007). A genomic background based method for association analysis in related individuals. *PLoS One* 2, e1274. doi: 10.1371/journal.pone.0001274
- Atkinson, O. (2013). Practical and effective management of foot lameness in dairy herds. *In Practice* 35, 171–182. doi: 10.1136/inp.f2014
- Becker, D. J., and Lowe, J. B. (2003). Fucose: biosynthesis and biological function in mammals. *Glycobiology* 13, 41–53. doi: 10.1093/glycob/cwg054
- Bicalho, R. C., Machado, V. S., and Caixeta, L. S. (2009). Lameness in dairy cattle: a debilitating disease or a disease of debilitated cattle? A cross-sectional study of lameness prevalence and thickness of the digital cushion. *J. Dairy Sci.* 92, 3175–3184. doi: 10.3168/jds.2008-1827
- Bicalho, R. C., and Oikonomou, G. (2013). Control and prevention of lameness associated with claw lesions in dairy cows. *Liv. Sci.* 156, 96–105. doi: 10.1016/j.livsci.2013.06.007
- Biffani, S., Pausch, H., Schwarzenbacher, H., and Biscarini, F. (2017). The effect of mislabeled phenotypic status on the identification of mutation-carriers from SNP genotypes in dairy cattle. *BMC Res. Notes* 10, 230. doi: 10.1186/s13104-017-2540-x
- Boda, E., Viganò, F., Rosa, P., Fumagalli, M., Labat-Gest, V., Tempia, F., et al. (2011). The GPR17 receptor in NG2 expressing cells: focus on in vivo maturation and participation in acute trauma and chronic damage. *Glia* 59, 1958–1973. doi: 10.1002/glia.21237
- Boettcher, P. J., Dekkers, J. C., Warnick, L. D., and Wells, S. J. (1998). Genetic analysis of clinical lameness in dairy cattle. *J. Dairy Sci.* 81, 1148–1156. doi: 10.3168/jds.S0022-0302(98)75677-2
- Buitenhuis, A. J., Lund, M. S., Thomasen, J. R., Thomsen, B., Nielsen, V. H., Bendixen, C., et al. (2007). Detection of quantitative trait loci affecting lameness and leg conformation traits in Danish Holstein cattle. *J. Dairy Sci.* 90, 472–481. doi: 10.3168/jds.S0022-0302(07)72649-8
- Buyse, S., Yang, G., Matise, T. C., and Gordon, D. (2009). When a case is not a case: effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Hum. Heredity* 67, 287–292. doi: 10.1159/000194981
- Cebamanos, L., Gray, A., Stewart, I., and Tenesa, A. (2014). Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. *Bioinformatics* 30, 1177–1179. doi: 10.1093/bioinformatics/btt754

ACKNOWLEDGMENTS

Bethany Griffiths, Lara Robinson, Nick Britten, Hannah Tatham, Rebecca Jenkin, and Nikos Kakatsidis are gratefully acknowledged for their assistance during data collection. Dr George Wiggans is gratefully acknowledged for converting the Affymetrix genotypes into an Illumina chip format.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00926/full#supplementary-material>

SUPPLEMENTARY FIGURE 1 | Genome-wide association (top) and regional heritability mapping (bottom) results by traits: Figure shows the Manhattan plots for each one of the traits, where concordant regions between both analyses were observed (A) digital dermatitis for farm records; (B) interdigital hyperplasia for farm records; (C) interdigital hyperplasia for combined records and; (D) sole haemorrhage for combined records. Red lines correspond to the genome-wide significant threshold, and blue lines correspond to the genome-wide suggestive threshold.

- Cha, E., Hertl, J. A., Bar, D., and Grohn, Y. T. (2010). The cost of different types of lameness in dairy cows calculated by dynamic programming. *Prev. Vet. Med.* 97, 1–8. doi: 10.1016/j.prevetmed.2010.07.011
- Egger-Danner, C., Nielsen, P., Fiedler, A., Müller, K., Fjeldaa, T., Döpfer, D., et al. (2015). *Icar claw health atlas*, [Online]. Available: https://www.icar.org/Documents/ICAR_Claw_Health_Atlas.pdf [Accessed September 7, 2018].
- Gernand, E., Rehbein, P., Von Borstel, U. U., and König, S. (2012). Incidences of and genetic parameters for mastitis, claw disorders, and common health traits recorded in dairy cattle contract herds. *J. Dairy Sci.* 95, 2144–2156. doi: 10.3168/jds.2011-4812
- Gilmour, A. R., Gogel, B., Cullis, B. R., and Thompson, R., (2009). *2009 ASReml user guide release 3.0*. Hemel Hempstead: VSN International Ltd.
- Green, L. E., Hedges, V. J., Schukken, Y. H., Blowey, R. W., and Packington, A. J. (2002). The impact of clinical lameness on the milk yield of dairy cows. *J. Dairy Sci.* 85, 2250–2256. doi: 10.3168/jds.S0022-0302(02)74304-X
- Hamanaka, R. B., Nigdelioglu, R., Meliton, A. Y., Tian, Y., Witt, L. J., O'leary, E., et al. (2017). Inhibition of phosphoglycerate dehydrogenase attenuates bleomycin-induced pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 58, 585–593. doi: 10.1165/rcmb.2017-0186OC
- Heringstad, B., Egger-Danner, C., Charfeddine, N., Pryce, J. E., Stock, K. F., Kofler, J., et al. (2018). Invited review: genetics and claw health: opportunities to enhance claw health by genetic selection. *J. Dairy Sci.* 101, 4801–4821. doi: 10.3168/jds.2017-13531
- Hocher, B., Schwarz, A., Fagan, K. A., Thone-Reineke, C., El-Hag, K., Kusserow, H., et al. (2000). Pulmonary fibrosis and chronic lung inflammation in ET-1 transgenic mice. *Am. J. Respir. Cell Mol. Biol.* 23, 19–26. doi: 10.1165/ajrcmb.23.1.4030
- Holzhauser, M., Hardenberg, C., Bartels, C. J. M., and Frankena, K. (2006). Herd- and cow-level prevalence of digital dermatitis in the Netherlands and associated risk factors. *J. Dairy Sci.* 89, 580–588. doi: 10.3168/jds.S0022-0302(06)72121-X
- Huxley, J. N. (2013). Impact of lameness and claw lesions in cows on health and production. *Liv. Sci.* 156, 64–70. doi: 10.1016/j.livsci.2013.06.012
- Koenig, S., Sharifi, A. R., Wentrot, H., Landmann, D., Eise, M., and Simianer, H. (2005). Genetic parameters of claw and foot disorders estimated with logistic models. *J. Dairy Sci.* 88, 3316–3325. doi: 10.3168/jds.S0022-0302(05)73015-0
- Laursen, M. V., Boelling, D., and Mark, T. (2009). Genetic parameters for claw and leg health, foot and leg conformation, and locomotion in Danish Holsteins. *J. Dairy Sci.* 92, 1770–1777. doi: 10.3168/jds.2008-1388

- Lee, S. H., Wray, N. R., Goddard, M. E., and Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* 88, 294–305. doi: 10.1016/j.ajhg.2011.02.002
- Maekawa, A., Balestrieri, B., Austen, K. F., and Kanaoka, Y. (2009). GPR17 is a negative regulator of the cysteinyl leukotriene 1 receptor response to leukotriene D(4). *Proc. Natl. Acad. Sci. U. S. A.* 106, 11685–11690. doi: 10.1073/pnas.0905364106
- Malchiodi, F., Koeck, A., Mason, S., Christen, A. M., Kelton, D. F., Schenkel, F. S., et al. (2017). Genetic parameters for hoof health traits estimated with linear and threshold models using alternative cohorts. *J. Dairy Sci.* 100, 2828–2836. doi: 10.3168/jds.2016-11558
- Malchiodi, F., Schenkel, F. S., Christen, A. M., Kelton, D. F., and Miglior, F. (2018). “Genome-wide association study and functional analysis of infectious and horn type hoof lesions in Canadian Holstein cattle,” in *World Congress on Genetics Applied to Livestock Production*.
- Matsushima, H., Yamada, N., Matsue, H., and Shimada, S. (2004). The effects of endothelin-1 on degranulation, cytokine, and growth factor production by skin-derived mast cells. *Eur. J. Immunol.* 34, 1910–1919. doi: 10.1002/eji.200424912
- Oberbauer, A. M., Berry, S. L., Belanger, J. M., Mcgoldrick, R. M., Pinos-Rodriguez, J. M., and Famula, T. R. (2013). Determining the heritable component of dairy cattle foot lesions. *J. Dairy Sci.* 96, 605–613. doi: 10.3168/jds.2012-5485
- Oikonomou, G., Banos, G., Machado, V., Caixeta, L., and Bicalho, R. C. (2014). Short communication: genetic characterization of digital cushion thickness. *J. Dairy Sci.* 97, 532–536. doi: 10.3168/jds.2013-7212
- Olmos, G., Boyle, L., Horan, B., Berry, D. P., O’connor, P., Mee, J. F., et al. (2009). Effect of genetic group and feed system on locomotion score, clinical lameness and hoof disorders of pasture-based Holstein–Friesian cows. *Animal* 3, 96–107. doi: 10.1017/S1751731108003248
- Ossent, P., and Lischer, C. (1998). Bovine laminitis: the lesions and their pathogenesis. *In Practice* 20, 415–427. doi: 10.1136/inpract.20.8.415
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Rockey, D. C., and Chung, J. J. (1996). Endothelin antagonism in experimental hepatic fibrosis. Implications for endothelin in the pathogenesis of wound healing. *J. Clin. Invest.* 98, 1381–1388. doi: 10.1172/JCI118925
- Scholey, R. A., Blowey, R. W., Murray, R. D., Smith, R. F., Cameron, J., Massey, J. P., et al. (2012). Investigating host genetic factors in bovine digital dermatitis. *Vet. Rec.* 171, 624–625. doi: 10.1136/vr.101251
- Schopke, K., Gomez, A., Dunbar, K. A., Swalve, H. H., and Dopfer, D. (2015). Investigating the genetic background of bovine digital dermatitis using improved definitions of clinical status. *J. Dairy Sci.* 98, 8164–8174. doi: 10.3168/jds.2015-9485
- Smith, T. J. (2005). Insights into the role of fibroblasts in human autoimmune diseases. *Clin. Exp. Immunol.* 141, 388–397. doi: 10.1111/j.1365-2249.2005.02824.x
- Solano, L., Barkema, H. W., Mason, S., Pajor, E. A., Leblanc, S. J., and Orsel, K. (2016). Prevalence and distribution of foot lesions in dairy cattle in Alberta, Canada. *J. Dairy Sci.* 99, 6828–6841. doi: 10.3168/jds.2016-10941
- Swalve, H. H., Floren, C., Wensch-Dorendorf, M., Schopke, K., Pijl, R., Wimmers, K., et al. (2014). A study based on records taken at time of hoof trimming reveals a strong association between the IQ motif-containing GTPase-activating protein 1 (IQGAP1) gene and sole hemorrhage in Holstein cattle. *J. Dairy Sci.* 97, 507–519. doi: 10.3168/jds.2013-6997
- Van Der Linde, C., De Jong, G., Koenen, E. P., and Eding, H. (2010). Claw health index for Dutch dairy cattle based on claw trimming and conformation data. *J. Dairy Sci.* 93, 4883–4891. doi: 10.3168/jds.2010-3183
- Van Der Spek, D., Van Arendonk, J. A., Vallee, A. A., and Bovenhuis, H. (2013). Genetic parameters for claw disorders and the effect of preselecting cows for trimming. *J. Dairy Sci.* 96, 6070–6078. doi: 10.3168/jds.2013-6833
- Van Der Waaij, E. H., Holzhauer, M., Ellen, E., Kamphuis, C., and De Jong, G. (2005). Genetic parameters for claw disorders in Dutch dairy cattle and correlations with conformation traits. *J. Dairy Sci.* 88, 3672–3678. doi: 10.3168/jds.S0022-0302(05)73053-8
- Vancheeswaran, R., Azam, A., Black, C., and Dashwood, M. R. (1994). Localization of endothelin-1 and its binding sites in scleroderma skin. *J. Rheumatol.* 21, 1268–1276.
- Wiltshire, S. A., Leiva-Torres, G. A., and Vidal, S. M. (2011). Quantitative trait locus analysis, pathway analysis, and consomic mapping show genetic variants of Tnni3k, Fpgt, or H28 control susceptibility to viral myocarditis. *J. Immunol.* 186, 6398–6405. doi: 10.4049/jimmunol.1100159
- Xu, S. (2003). Theoretical basis of the Beavis effect. *Genetics* 165, 2259–2268.
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310
- Zwald, N. R., Weigel, K. A., Chang, Y. M., Welper, R. D., and Clay, J. S. (2004). Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values. *J. Dairy Sci.* 87, 4287–4294. doi: 10.3168/jds.S0022-0302(04)73573-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sánchez-Molano, Bay, Smith, Oikonomou and Banos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership