The generation of the largest to date pig gut metagenomic dataset and the insights gained into microbiome composition and function

by

DANIELA GAIO

A Thesis Submitted In
Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

under the supervision of
AARON E. DARLING

Faculty of Science
University of Technology Sydney
Sydney, NSW, Australia

February 2021

to my father

Certificate of Original Authorship

I, Daniela Gaio, declare that this thesis is submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Daniela Gaio

Production Note:
Signature removed
prior to publication.

Date: 2021-02-09

Acknowledgements

There is this common belief that the PhD journey is by default drenched in suffering. Well, it wasn't for me, and I believe I have mostly to thank my principal supervisor Aaron for it. I want to thank you for your time, your patience, the positive reinforcement, and your many teachings. I am proud of having landed in the bioinformatics territory and I am determined to continue. Without your support it would have simply not been as good.

Thank you Merlino, for your unconditional love for the past two and more demanding years. Thank you Merlino and Semola, for making my every day better.

Thank you Shima Baradaran, Maggie Cheng, Shannon Hawkins, Cath Burke, Ausgem, and the ithree Institute for allowing me to conduct this work and providing the necessary support to do it. Thank you Vincent Ruffles for the chats, the encouragement, for being my number one supporter in my Stage 3. Thank you Mike Lake for your help while I was getting acquainted with the HPC.

I would like to thank my family and my mother, that in a way or another remind me now and then that despite my constant distance I am far from alone. In particular, thank you Maurizio, Sara, zio Livio, zia Maria Vittoria & zio Domenico. A thank you to my extended family; thank you Silvia for loving my father, for freeing me of lots of other worries, for being nice to me.

I would like to thank the precious souls that I met on this journey, that improved the quality of my time, that taught me something: Martin Muggli, Beatriz Garcia Jimenez, Cameron Reid, Giulia Ballerin, Michael Widjaja, Jos van Nijnatten, Barbara Brito, Fred Jaya, Nehleh Kargarfard, Hesham Wahid Badr, Daniel Jakobs, Stephanie Gonzales, Javi Salinas,

Mauricio Sierra, Nico Soruco, Ekaitz Santamaria, Siward van Faassen, Joost Oosterhof, Emma Gardiner, Ilse Dekker, Adam Pople, Gianfranco, Daniel Tamas, Michal Kara, Alex Khun, Nanami Suzuki, Eduardo Aurelio Salomon Diaz, Milenko Brzovic, Antonio Rodriguez, Cezar Dounis Accioly, Duilio Santovito, Paloma Ve, all the cool people from Bellevue, and the So Many Flies In The Office squad, for adding some office-silliness during this journey. Thank you Tiziana Zingali for the numerous laughs and for being my partner in crime during the animal trial, and Famheda Abdi, Ahmed Moustafa, Edward Ni for your company during covid time. Thank you Milagro Elstak, Joanna Keogh, Florian Kleinknecht, Thera en familie, die me waarderen en ondersteunen in mijn advonturen in de wijde wereld, ondanks de afstand. Dank je Steve, dat je in mijn hoofd blijft zweven. Thank you Adam Farzaneh for visiting me and for all the chats. Thank you Simone Arnold for the hours and hours of monologues correspondence; Mauro Corda for the surf sessions in good company; Giovanni Piredda for the casual nihilistic exchange; and Fabiano Cuccu for the river dips and not just. Thank you Giuliano Serra, Laura Mura, Gabriele Cau for always being there; and Adrien Loustanau for your support and the sense of comfort during these last months, well…years.

And of course thank you dad for making me feel alive, for making me giggle in self-talk, for having taught me that everything is and is not, and it is all beautiful nonetheless.

Daniela Gaio

Sydney, Australia, 2021

Thesis format

A thesis by compilation.

Publications arising

1. Gaio, D., To, J., Liu, M., Monahan, L., Anantanawat, K., & Darling, A. E. (2019). Hackflex: low cost Illumina sequencing library construction for high sample counts. *BioRxiv*, 779215.
   doi: https://doi.org/10.1101/779215
   Current status: under peer-review; available as a pre-print
2. Gaio D., DeMaere Z. M., Anatanawat K., Eamens G. J., Liu, M., Zingali T., Falconer L., Chapman T. A., Djordjevic S. P., & Darling, A. E. (2020). Community composition and development of the post-weaning piglet gut microbiome. *BioRxiv*, 211326.
   doi: https://doi.org/10.1101/2020.07.20.211326
   Current status: under peer-review; available as a pre-print
3. Gaio D., DeMaere Z. M., Anatanawat K., Chapman T. A., Djordjevic S. P., & Darling, A. E. (2020). Post-weaning shifts in microbiome composition and metabolism revealed by over 25,000 pig gut metagenome assembled genomes. *BioRxiv*, 253872.
   doi: https://doi.org/10.1101/2020.08.17.253872
   Current status: under peer-review; available as a pre-print.

Contents

List of Tables

List of Figures

List of Supplementary Tables

List of Supplementary Figures

Abstract

Early weaning and intensive farming practices predispose piglets to the development of infectious and often lethal diseases, against which antibiotics are used. Besides contributing to the build-up of antimicrobial resistance, antibiotics are known to substantially alter the gut microbial composition. While extensive literature reports on the dysbiotic effects of orally administered antibiotics, fewer studies established the effects of intramuscular antibiotic treatment on the gut community, and none, to our knowledge, studied the effects of intramuscular neomycin.

In order to study the effects of intramuscular neomycin, and of two probiotic formulations, we sampled 911 faecal microbial communities from 126 post-weaning piglets for the duration of 5 weeks.

The processing of such a large amount of samples required the optimization of costs with the maintenance of sample quality. To this end, we developed a sequencing library preparation technique, called Hackflex, which allows the generation of Illumina-compatible sequencing libraries with up to an 11-fold cost reduction. The sequencing generated the largest piglet gut microbiome dataset up to date, comprising of 8Tb of metagenomic shotgun sequence data, corresponding to 27 billion read pairs.

A phylogenetic framework allowed us to interrogate the data. We found factors such as breed, litter and tiny differences in age to be associated with differences in the microbiome and we determined the persistence of those correlations. Specific microbial compositional changes were found to correlate with the probiotic and the antibiotic treatments.

Lastly, reads were pooled and co-assembled by host, and the differential coverage abundance information was used in the process of binning, to produce metagenomic assembled genomes (MAGs). This method constructs draft genomes without the need for reference databases, which are unequivocally biased towards better-represented organisms. Clustered MAGs allowed us to track microbial shifts at the species- and strain-level occurring between the 3rd and the 9th week of life of post-weaning piglets, and the functional analysis on the carbohydrate enzyme repertoire substantially expanded the current knowledge of species-to-enzyme mapping.

The contribution of this work to the current knowledge of the post-weaning gut microbial development opens the doors to the possibility to optimise the chances of survival of specific strains by knowing exactly when they should be administered; and to predict a strain's substrate utilization and in such way, engineer probiotic and prebiotic cocktails for the improvement of substrate energy efficiency. These possibilities are applicable to the human (*e.g.* diet) as well as the livestock settings (*e.g*. diet, methane emission control).