

# **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

© 2021

Daniela Gaio

All Rights Reserved

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 itthree 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 waarden en ondersteunen in mijn avonturen 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

<b>CERTIFICATE OF ORIGINAL AUTHORSHIP .....</b>	<b>III</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>IV</b>
<b>THESIS FORMAT .....</b>	<b>VI</b>
<b>PUBLICATIONS ARISING .....</b>	<b>VII</b>
<b>LIST OF TABLES .....</b>	<b>XII</b>
<b>LIST OF FIGURES.....</b>	<b>XII</b>
<b>LIST OF SUPPLEMENTARY TABLES.....</b>	<b>XIII</b>
<b>LIST OF SUPPLEMENTARY FIGURES.....</b>	<b>XIII</b>
<b>ABSTRACT.....</b>	<b>XVI</b>
<b>CHAPTER 1 .....</b>	<b>1</b>
<b>INTRODUCTION .....</b>	<b>1</b>
<b>1.1 PATHOGENS AND ANTIMICROBIAL RESISTANCE IN LIVESTOCK.....</b>	<b>1</b>
1.1.1 <i>Disease susceptibility in post-weaning piglets.....</i>	<i>1</i>
1.1.2 <i>Antibiotic use.....</i>	<i>2</i>
1.1.3 <i>Rise and transmission of antimicrobial resistance (AMR).....</i>	<i>3</i>
1.1.4 <i>AMR gene transfer.....</i>	<i>4</i>
1.1.5 <i>Prophage induction.....</i>	<i>6</i>
1.1.6 <i>AMR transmission among animals.....</i>	<i>7</i>
1.1.7 <i>AMR transmission to the environment and to humans.....</i>	<i>8</i>
<b>1.2 THE MICROBIOME .....</b>	<b>10</b>
1.2.1 <i>The gut microbiome.....</i>	<i>10</i>
1.2.2 <i>Dynamicity, stability, redundancy and dysbiosis .....</i>	<i>11</i>
1.2.3 <i>Roles of the gut microbiota .....</i>	<i>13</i>
1.2.4 <i>The microbiome of the pig.....</i>	<i>16</i>
1.2.5 <i>Antibiotic-induced dysbiosis.....</i>	<i>18</i>
<b>1.3 MICROBIOME MODULATING STRATEGIES AGAINST ENTERIC DISEASES.....</b>	<b>21</b>
1.3.1 <i>Non-probiotic strategies.....</i>	<i>21</i>
1.3.2 <i>Probiotic strategies.....</i>	<i>24</i>
<b>1.4 METAGENOMICS .....</b>	<b>30</b>
1.4.1 <i>Metagenomics (not the 16S rRNA type!).....</i>	<i>30</i>
1.4.2 <i>Thousand and one sources of artefact and bias, and the precautions we took.....</i>	<i>31</i>
1.4.3 <i>Assembly.....</i>	<i>33</i>
1.4.4 <i>Binning.....</i>	<i>35</i>
1.4.5 <i>Data analysis.....</i>	<i>38</i>
<b>1.5 RESEARCH AIM .....</b>	<b>39</b>
<b>1.6 OUTLINE OF THESIS.....</b>	<b>39</b>
1.6.1 <i>Objectives.....</i>	<i>39</i>
1.6.2 <i>Chapter Summaries.....</i>	<i>40</i>
<b>1.7 REFERENCES.....</b>	<b>42</b>

<b>CHAPTER 2 .....</b>	<b>54</b>
<b>HACKFLEX: LOW COST ILLUMINA SEQUENCING LIBRARY CONSTRUCTION FOR HIGH SAMPLE COUNTS .....</b>	<b>54</b>
2.1 AUTHORSHIP DECLARATION .....	54
2.2 ABSTRACT .....	55
2.3 INTRODUCTION .....	55
2.4 METHODS .....	57
2.4.1 <i>Genomic DNA preparation</i> .....	57
2.4.2 <i>Barcode design</i> .....	57
2.4.3 <i>Nextera Flex sequencing libraries preparation</i> .....	58
2.4.4 <i>Tagmentation and Hackflex sequencing library preparation</i> .....	59
2.4.5 <i>Preparation of additional Illumina libraries</i> .....	60
2.4.6 <i>Nanopore library preparation and sequencing</i> .....	61
2.4.7 <i>Data analysis</i> .....	61
2.4.8 <i>Barcode demultiplexing</i> .....	61
2.4.9 <i>Processing of libraries before mapping</i> .....	61
2.4.10 <i>Hybrid assembly of Nanopore and Illumina data</i> .....	62
2.4.11 <i>Short read mapping and coverage analysis</i> .....	63
2.4.12 <i>GC content</i> .....	63
2.4.13 <i>Data availability</i> .....	64
2.5 RESULTS .....	64
2.5.1 <i>Hackflex library preparation and sequencing</i> .....	64
2.5.2 <i>Barcode distribution and quality</i> .....	64
2.5.3 <i>Quality of raw reads</i> .....	66
2.5.4 <i>Cleaning</i> .....	67
2.5.5 <i>Fragment size</i> .....	68
2.5.6 <i>Coverage</i> .....	69
2.5.7 <i>GC content</i> .....	71
2.5.8 <i>Low coverage regions</i> .....	72
2.5.9 <i>Additional libraries: yield and GC coverage bias</i> .....	72
2.6 DISCUSSION .....	73
2.6.1 <i>Caveats and limitations</i> .....	73
2.7 CONCLUSION .....	74
2.8 CONFLICTS OF INTEREST .....	75
2.9 FUNDING INFORMATION .....	75
2.10 AUTHOR CONTRIBUTIONS .....	75
2.11 SUPPLEMENTARY FILES .....	75
2.12 REFERENCES .....	76
2.13 SUPPLEMENTARY MATERIAL .....	77
<b>CHAPTER 3 .....</b>	<b>87</b>
<b>COMMUNITY COMPOSITION AND DEVELOPMENT OF THE POST- WEANING PIGLET GUT MICROBIOME .....</b>	<b>87</b>
3.1 AUTHORSHIP DECLARATION .....	87
3.2 ABSTRACT .....	88
3.3 INTRODUCTION .....	88
3.4 METHODS .....	90
3.4.1 <i>Pig Trial</i> .....	90
3.4.2 <i>Faecal sample processing</i> .....	92
3.4.3 <i>Positive controls</i> .....	92
3.4.4 <i>DNA extraction</i> .....	93
3.4.5 <i>Library preparation</i> .....	94
3.4.6 <i>Size selection and purification</i> .....	94

3.4.7	<i>Normalization and sequencing</i>	95
3.4.8	<i>Sequence data processing</i>	95
3.4.9	<i>Determination of microbial diversity among samples</i>	95
3.4.10	<i>Batch effects</i>	96
3.4.11	<i>Technical analysis of positive controls</i>	97
3.4.12	<i>Data availability</i>	97
3.5	RESULTS	97
3.5.1	<i>Comparison of the expected and the observed taxonomic profile of the positive controls</i>	97
3.5.2	<i>Phylogenetic diversity of positive controls and how it compares to their taxonomic profile</i>	99
3.5.3	<i>Phylogenetic diversity of piglet gut communities immediately after weaning</i>	100
3.5.4	<i>Maternal effect on phylogenetic diversity</i>	102
3.5.5	<i>A strong effect of aging on phylogenetic diversity</i>	102
3.5.6	<i>Effect of antibiotic and probiotic treatment on alpha diversity</i>	107
3.5.7	<i>Effect of antibiotic and probiotic treatment on beta diversity</i>	109
3.5.8	<i>Association between weight and community composition</i>	109
3.5.9	<i>Effect of treatments on weight gain</i>	110
3.6	DISCUSSION	110
3.6.1	<i>Technical controls in metagenomic studies and methodological limitations</i>	114
3.7	CONCLUSION	115
3.8	CONFLICTS OF INTEREST	116
3.9	FUNDING INFORMATION	116
3.10	AUTHOR CONTRIBUTIONS	116
3.11	ACKNOWLEDGEMENTS	117
3.12	SUPPLEMENTARY FILES	117
3.13	REFERENCES	118
3.14	SUPPLEMENTARY MATERIAL	124
	<b>CHAPTER 4</b>	<b>158</b>
	<b>POST-WEANING SHIFTS IN MICROBIOME COMPOSITION AND METABOLISM REVEALED BY OVER 25,000 PIG GUT METAGENOME ASSEMBLED GENOMES</b>	<b>158</b>
4.1	AUTHORSHIP DECLARATION	158
4.2	ABSTRACT	159
4.3	INTRODUCTION	159
4.4	METHODS	161
4.4.1	<i>Metagenomic samples</i>	161
4.4.2	<i>Sequence data processing</i>	161
4.4.3	<i>Pooled assembly and binning</i>	162
4.4.4	<i>MAGs quality assignment</i>	162
4.4.5	<i>MAGs taxonomic assignment</i>	162
4.4.6	<i>MAGs dereplication</i>	163
4.4.7	<i>Comparison of dRep clusters with GTDB- taxonomic assignments</i>	163
4.4.8	<i>Data analysis</i>	163
4.4.9	<i>Differential abundance analysis</i>	164
4.4.10	<i>Gene function analysis</i>	165
4.4.11	<i>Analysis workflow and scripts</i>	165
4.5	RESULTS	166
4.5.1	<i>Completeness and contamination of MAGs</i>	167

4.5.2	<i>MAG ANI clusters and taxonomic clusters</i> .....	168
4.5.3	<i>High-level taxonomy of the piglet gut</i> .....	169
4.5.4	<i>Time trend</i> .....	169
4.5.5	<i>A remarkable tightly regulated compositional shift</i> .....	170
4.5.6	<i>Diversity</i> .....	174
4.5.7	<i>Effects of age, breed, cohousing, and weight</i> .....	175
4.5.8	<i>The predicted proteome of the piglet</i> .....	176
4.5.9	<i>CAZy species-specificity and their distribution across the pig population</i> .....	177
4.5.10	<i>The piglet carbohydrate proteome across time</i> .....	179
4.6	DISCUSSION.....	180
4.6.1	<i>More MAGs were obtained from more frequently sampled subjects</i> 181	
4.6.2	<i>Concordance of taxonomic MAG clusters with ANI clusters</i> .....	181
4.6.3	<i>Time dictates composition, down to the species level</i> .....	181
4.6.4	<i>Strain differences strongly associated with the timeline of gut development</i> .....	183
4.6.5	<i>Possible drivers of a tight regulation: immune system versus diet</i> 184	
4.6.6	<i>Probiotic strains go missing: did they fail to assemble or fail to colonize?</i> .....	185
4.6.7	<i>Assessment of probiotic treatment effect on other taxa</i> .....	186
4.6.8	<i>Carbohydrate metabolism based on predicted protein CAZy profile</i> 187	
4.6.9	<i>Limitations</i> .....	188
4.7	CONFLICTS OF INTEREST.....	189
4.8	FUNDING INFORMATION.....	190
4.9	AUTHOR CONTRIBUTIONS.....	190
4.10	ACKNOWLEDGEMENTS.....	190
4.11	SUPPLEMENTARY FILES.....	190
4.12	REFERENCES.....	191
4.13	SUPPLEMENTARY MATERIAL.....	196
<b>5</b>	<b>CHAPTER 5</b> .....	<b>231</b>
<b>5</b>	<b>GENERAL DISCUSSION &amp; FUTURE DIRECTIONS</b> .....	<b>231</b>
5.1	DISCUSSION.....	231
5.1.1	<i>Were the objectives met?</i> .....	231
5.1.2	<i>Other lessons learnt</i> .....	233
5.1.3	<i>Applications to the industry</i> .....	235
5.2	FUTURE DIRECTIONS.....	235
5.2.1	<i>Tracking the journeys of AMR genes</i> .....	235
5.2.2	<i>Further exploration of the dataset</i> .....	236
5.3	REFERENCES.....	238

# List of Tables

TABLE 2.1. OVERVIEW OF LIBRARY CONSTRUCTION METHODS AND READS OBTAINED. ....	62
TABLE 2.2. COVERAGE OBTAINED FROM MAPPING LIBRARIES AGAINST ALL CONTIGS OF THE <i>E. COLI</i> MG1655 NANOPORE ASSEMBLY.....	68

# List of Figures

FIGURE 1.1. SUMMARY OF MECHANISMS OF AMR TRANSFER AND RETAINMENT. ....	6
FIGURE 1.2. BINNING WITH THE USE OF DIFFERENTIAL ABUNDANCE PATTERNS.....	38
FIGURE 2.1. PRICE PER SAMPLE (IN AUD) WHEN PROCESSING UP TO 96x50 (4800) SAMPLES. ...	56
FIGURE 2.2. UNIQUE BARCODE DISTRIBUTION ACROSS THE 96 HACKFLEX LIBRARIES.....	66
FIGURE 2.3. QUALITY (PHRED) SCORES FOR READS FROM STANDARD FLEX, 1:50 FLEX AND HACKFLEX LIBRARY. ....	67
FIGURE 2.4. FRAGMENT SIZE DISTRIBUTION OF STANDARD FLEX, 1:50 FLEX AND HACKFLEX.....	69
FIGURE 2.5. COVERAGE DISTRIBUTION OF UNIQUE READS ACROSS THE <i>E. COLI</i> MG1655 GENOME. .....	70
FIGURE 2.6. GC CONTENT BIAS. ....	71
FIGURE 3.1. TIMELINE. ....	92
FIGURE 3.2. SIGNIFICANT CORRELATIONS OF PHYLOGENETIC DIVERSITY WITH SPECIFIED FACTORS. .....	101
FIGURE 3.3. EFFECT OF TIME ON BETA DIVERSITY. ....	104
FIGURE 3.4. ABUNDANCE HEATMAP. ....	107
FIGURE 4.1. MAGS OBTAINED VERSUS NUMBER OF TIME POINT SAMPLES AVAILABLE PER SUBJECT. THE NUMBER OF MAGS OBTAINED PER SUBJECT DOUBLED WITH THE NUMBER OF TIME POINT SAMPLES AVAILABLE PER SUBJECT. THE GAINS OF MAGS PER SUBJECT STARTED TO DIMINISH AFTER 4 TIME POINT SAMPLES PER SUBJECT.....	166
FIGURE 4.2. QUALITY OF MAGS. ....	167
FIGURE 4.3. TEMPORAL SHIFT OF SAMPLES FROM ANI CLUSTERED MAGS AND GTDB TAXONOMICALLY CLUSTERED MAGS. ....	170
FIGURE 4.4. THE TWENTY-SIX MOST ABUNDANT TAXA IN THE PIGLET POPULATION AND THEIR RELATIVE ABUNDANCE PROFILE OVER TIME.....	173
FIGURE 4.5. SIGNIFICANTLY CHANGING GTDB-SPECIES WITH TIME. ....	174
FIGURE 4.6. CAZY ENZYMES IN THE PIG POPULATION AND THEIR SPECIES REPRESENTATION.....	179

# List of Supplementary Tables

SUPPLEMENTARY TABLE 2.1. QUANTITIES AND COSTS OF HACKFLEX REAGENTS AS PER RETAIL PRICE. ....	77
SUPPLEMENTARY TABLE 2.2. SUMMARY OF PROTOCOL DETAILS FOR THE CONSTRUCTION OF THE ADDITIONAL LIBRARIES AND OBTAINED YIELDS. ....	78
SUPPLEMENTARY TABLE 2.3. DETAILED COVERAGE SUMMARY. ....	83
SUPPLEMENTARY TABLE 2.4. GENOMIC POSITIONS OF SITES WITH LOWEST COVERAGE FOR STANDARD FLEX, 1:50 FLEX AND HACKFLEX. ....	86

# List of Supplementary Figures

SUPPLEMENTARY FIGURE 2.1. SCHEMATIC OVERVIEW OF DATA ANALYSIS METHODS USED IN THIS STUDY. ....	79
SUPPLEMENTARY FIGURE 2.2. COVERAGE DISTRIBUTION OF UNIQUE READS ACROSS THE E. COLI MG1655 GENOME FOR EACH OF THE ASSEMBLIES. ....	80
SUPPLEMENTARY FIGURE 2.3. BARCODE INDEXES PERFORMANCE. ....	81
SUPPLEMENTARY FIGURE 2.4. BARCODE GC CONTENT. ....	82
SUPPLEMENTARY FIGURE 2.5. LOWEST COVERAGE SITES FOR STANDARD FLEX, 1:50 FLEX AND HACKFLEX. ....	84
SUPPLEMENTARY FIGURE 2.6. GC COVERAGE BIAS FOR ADDITIONAL LIBRARIES. ....	85
SUPPLEMENTARY FIGURE 3.1. PIGLETS PLACEMENTS ACROSS ROOMS AND PENS. ....	124
SUPPLEMENTARY FIGURE 3.2. SAMPLE COLLECTION. ....	125
SUPPLEMENTARY FIGURE 3.3. READ COUNTS DISTRIBUTION. ....	126
SUPPLEMENTARY FIGURE 3.4. SUMMARISED WORKFLOW. ....	127
SUPPLEMENTARY FIGURE 3.5. PRINCIPAL COMPONENT ANALYSIS OF BATCH EFFECT BEFORE BATCH EFFECT REMOVAL. ....	128
SUPPLEMENTARY FIGURE 3.6. BATCH EFFECT ON ALPHA AND BETA DIVERSITY BEFORE BATCH EFFECT REMOVAL. ....	129
SUPPLEMENTARY FIGURE 3.7. BATCH EFFECT ON ALPHA AND BETA DIVERSITY AFTER BATCH EFFECT REMOVAL. ....	130
SUPPLEMENTARY FIGURE 3.8. TAXONOMIC ASSIGNMENT OF READS FROM POSITIVE CONTROL SAMPLES. ....	131
SUPPLEMENTARY FIGURE 3.9. EXPECTED AND OBSERVED RELATIVE ABUNDANCE OF MOCK COMMUNITY MEMBERS. ....	132
SUPPLEMENTARY FIGURE 3.10. CONTAMINATION RATE IN MOCK COMMUNITY SAMPLES. ....	133
SUPPLEMENTARY FIGURE 3.11. CONTAMINATION RATE IN D-SCOUR™ SAMPLES. ....	134
SUPPLEMENTARY FIGURE 3.12. CONTAMINATION RATE IN COLIGUARD® SAMPLES. ....	135
SUPPLEMENTARY FIGURE 3.13. ALPHA PHYLOGENETIC DIVERSITY OF COHORTS. ....	136
SUPPLEMENTARY FIGURE 3.14. ALPHA PHYLOGENETIC DIVERSITY BY AGE GROUP. ....	137
SUPPLEMENTARY FIGURE 3.15. ALPHA PHYLOGENETIC DIVERSITY BY AGE BREED. ....	138
SUPPLEMENTARY FIGURE 3.16. UNROOTED ALPHA PHYLOGENETIC DIVERSITY BY MATERNAL AND NURSE SOW. ....	139
SUPPLEMENTARY FIGURE 3.17. BALANCE WEIGHTED ALPHA PHYLOGENETIC DIVERSITY BY MATERNAL AND NURSE SOW. ....	140
SUPPLEMENTARY FIGURE 3.18. PRINCIPAL COMPONENT ANALYSIS OF PIGLET SAMPLES BY NURSE SOW. ....	141
SUPPLEMENTARY FIGURE 3.19. PRINCIPAL COMPONENT ANALYSIS OF PIGLET SAMPLES BY MATERNAL SOW. ....	142
SUPPLEMENTARY FIGURE 3.20. TIME EFFECT ON BETA DIVERSITY PER COHORT. ....	143
SUPPLEMENTARY FIGURE 3.21. TIME EFFECT ON BETA DIVERSITY. ....	144

SUPPLEMENTARY FIGURE 3.22. LINEAGES DISPLAYING THE HIGHEST VARIATION IN BETA DIVERSITY ACROSS TIME. ....	145
SUPPLEMENTARY FIGURE 3.23. TIME TREND OF ALPHA DIVERSITY BY COHORT. ....	146
SUPPLEMENTARY FIGURE 3.24. ALPHA DIVERSITY BY COHORT ACROSS FOUR TIME POINTS. ....	147
SUPPLEMENTARY FIGURE 3.25. DELTAS OF UNROOTED ALPHA PHYLOGENETIC DIVERSITY BETWEEN TIME POINTS.....	148
SUPPLEMENTARY FIGURE 3.26. DELTAS OF BALANCE WEIGHTED ALPHA PHYLOGENETIC DIVERSITY BETWEEN TIME POINTS.....	149
SUPPLEMENTARY FIGURE 3.27. SIGNIFICANT DIFFERENCES IN BETA DIVERSITY BETWEEN COHORTS AT SPECIFIC TIME POINTS. ....	150
SUPPLEMENTARY FIGURE 3.28. CORRELATION OF WEIGHT WITH TAXONOMY AT TIME POINT 0. ....	151
SUPPLEMENTARY FIGURE 3.29. CORRELATION OF WEIGHT WITH TAXONOMY AT TIME POINT 2. ....	152
SUPPLEMENTARY FIGURE 3.30. CORRELATION OF WEIGHT WITH TAXONOMY AT TIME POINT 4. ....	153
SUPPLEMENTARY FIGURE 3.31. CORRELATION OF WEIGHT WITH TAXONOMY AT TIME POINT 6. ....	154
SUPPLEMENTARY FIGURE 3.32. CORRELATION OF WEIGHT WITH TAXONOMY AT TIME POINT 8. ....	155
SUPPLEMENTARY FIGURE 3.33. CORRELATION OF WEIGHT WITH TAXONOMY AT TIME POINT 10.....	156
SUPPLEMENTARY FIGURE 3.34. WEIGHT GAIN OF PIGLETS BY COHORT ACROSS THE TRIAL. ....	157
SUPPLEMENTARY FIGURE 4.1. DISTRIBUTION OF BINS OVER SUBJECTS AND COHORTS. ....	196
SUPPLEMENTARY FIGURE 4.2. WORKFLOW OF SEQUENCING DATA PROCESSING AND ANALYSIS. .	197
SUPPLEMENTARY FIGURE 4.3. DISTRIBUTION OF SCAFFOLD N50 ACROSS NEARLY COMPLETE MAGs.....	198
SUPPLEMENTARY FIGURE 4.4. THE FRACTION OF MAGs THAT ARE UNIQUE TO INDIVIDUAL SUBJECTS WHEN CLUSTERED AT 95% AND 99% ANI.....	199
SUPPLEMENTARY FIGURE 4.5. EXTENT OF AGREEMENT BETWEEN dREP ANI CLUSTERING AND GTDB TAXONOMIC CLUSTERING. ....	200
SUPPLEMENTARY FIGURE 4.6. TAXONOMIC PROFILE OF THE POSITIVE CONTROLS. ....	201
SUPPLEMENTARY FIGURE 4.7. TAXONOMIC COMPOSITION OF THE PIGLET GUT.....	202
SUPPLEMENTARY FIGURE 4.8. TEMPORAL SHIFT OF MAGs FROM PHYLOSEQ ANALYSIS. ....	203
SUPPLEMENTARY FIGURE 4.9. TEMPORAL SHIFT OF SAMPLES FROM CHECKM CLUSTERED MAGs, dREP- ANI CLUSTERED MAGs, AND GTDB CLUSTERED MAGs.....	204
SUPPLEMENTARY FIGURE 4.10. THE MOST ABUNDANT TAXA IN THE PIGLET POPULATION AND THEIR RELATIVE ABUNDANCE PROFILE OVER TIME. ....	205
SUPPLEMENTARY FIGURE 4.11. NUMBER OF SIGNIFICANT ABUNDANCE SHIFTS PER TIME INTERVAL. ....	206
SUPPLEMENTARY FIGURE 4.12. SIGNIFICANTLY SHIFTING TAXA BETWEEN THE FIRST AND THE SECOND WEEK.....	207
SUPPLEMENTARY FIGURE 4.13. SIGNIFICANTLY SHIFTING TAXA BETWEEN THE SECOND AND THE THIRD WEEK.....	208
SUPPLEMENTARY FIGURE 4.14. SIGNIFICANTLY SHIFTING TAXA BETWEEN THE THIRD AND THE FOURTH WEEK.....	209
SUPPLEMENTARY FIGURE 4.15. SIGNIFICANTLY SHIFTING TAXA BETWEEN THE FOURTH AND THE FIFTH WEEK.....	210
SUPPLEMENTARY FIGURE 4.16. SIGNIFICANTLY SHIFTING TAXA BETWEEN THE FIFTH AND THE SIXTH WEEK.....	211
SUPPLEMENTARY FIGURE 4.17. SPECIES SIGNIFICANTLY SHIFTING IN ABUNDANCE AT ALL TIME INTERVALS.....	212
SUPPLEMENTARY FIGURE 4.18. SAMPLE DIVERSITY ACROSS TIME. ....	213
SUPPLEMENTARY FIGURE 4.19. AGE AND BREED CORRELATIONS WITH COMPOSITION. ....	214
SUPPLEMENTARY FIGURE 4.20. MILD TAXONOMIC DIFFERENCES BETWEEN AGE GROUPS.....	215
SUPPLEMENTARY FIGURE 4.21. CORRELATION OF WEIGHT WITH COMPOSITION.....	216

SUPPLEMENTARY FIGURE 4.22. SEQUENCE IDENTITY AGAINST THE CAZY DATABASE AND PROPORTIONS ACROSS PHYLA.....	217
SUPPLEMENTARY FIGURE 4.23. DIVERSITY OF GLYCOSIDE TRANSFERASES. ....	218
SUPPLEMENTARY FIGURE 4.24. AA AND CE CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION.....	219
SUPPLEMENTARY FIGURE 4.25. PL CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION. ....	220
SUPPLEMENTARY FIGURE 4.26. CBM CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION. ....	221
SUPPLEMENTARY FIGURE 4.27. CBM CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION (PART 2).....	222
SUPPLEMENTARY FIGURE 4.28. GT CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION. ....	223
SUPPLEMENTARY FIGURE 4.29. GT CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION (PART 2).....	224
SUPPLEMENTARY FIGURE 4.30. GT CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION (PART 3).....	225
SUPPLEMENTARY FIGURE 4.31. GH CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION. ....	226
SUPPLEMENTARY FIGURE 4.32. GH CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION (PART 2).....	227
SUPPLEMENTARY FIGURE 4.33. GH CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION (PART 3).....	228
SUPPLEMENTARY FIGURE 4.34. GH CLASS ENZYMES TIME TREND AND SPECIES REPRESENTATION (PART 4).....	229
SUPPLEMENTARY FIGURE 4.35. COHESIN AND S-LAYER HOMOLOGY (SLH) ENZYMES TIME TREND AND SPECIES REPRESENTATION. ....	230

# 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).