Reduced Representation Genotyping for Bacterial Identification, Discovery and Genomic Analysis

BERENICE TALAMANTES-BECERRA

DIPLOMA OF FOOD SCIENCE AND TECHNOLOGY – CENTRO DE ESTUDIOS TECNOLÓGICOS, INDUSTRIAL Y DE SERVICIOS NO. 100 - MEXICO

BACHELOR OF ENGINEERING, BIOTECHNOLOGY - INSTITUTO TECNÓLOGICO DE SONORA – MEXICO

MASTER IN INTERNATIONAL BUSINESS – INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE MONTERREY, CAMPUS GUADALAJARA - MEXICO

MASTER OF BUSINESS ADMINISTRATION – GRADUATE SCHOOL OF BUSINESS – PERU

INSTITUTE FOR APPLIED ECOLOGY

FACULTY OF SCIENCE AND TECHNOLOGY

UNIVERSITY OF CANBERRA, ACT, 2601, AUSTRALIA

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF DOCTOR OF PHILOSOPHY AT THE UNIVERSITY OF CANBERRA, AUSTRALIA

-SEPTEMBER 2019-

DEDICATION

Dedicado a mi familia, mis gatos, mis perros, y Colashina

Y a Jason, porque que sin ti, nada de esto fuera posible

ABSTRACT

Bacterial identification methods are important for medical, environmental, food and industrial microbiology. Current bacterial identification methods range from low resolution techniques such as biochemical testing and sequencing of the 16S rRNA gene to high-resolution methods such as whole genome sequencing. There are few options in between. To fill this gap, I applied a reduced-representation sequencing technique (DArTseq) for bacterial identification and typing to the field of microbiology, specifically medical microbiology and environmental microbiology. To analyse reduced-representation sequencing data, I developed a bioinformatics pipeline, Currito3.1 DNA Fragment Analysis Software for bacterial identification and strain typing. To meet these targets on medical and environmental microbiology, this thesis presents results from two case studies. The first case study involved genotyping 165 bacterial isolates previously identified using conventional methods, provided by the Microbiology Department of Canberra Public Hospital. These were processed with reduced-representation sequencing, using three combinations of restriction enzymes: *Pst*I with *Mse*I, *Pst*I with *Hpa*II and *Mse*I with *Hpa*II. All bacterial samples were correctly identified to genus and species by each of the three combinations of restriction enzymes. In the second case study, bacterial isolates were obtained from compost, domestic hot water systems and artesian bores of the Great Artesian Basin. The sampling locations represented extreme environments with temperatures as high as to 98°C. The study resulted in the isolation of 99 bacterial strains of the thermophilic genera *Anoxybacillus*, *Geobacillus* and *Parageobacillus*, from which 8 samples were selected for whole-genome sequencing. Identifications using reduced-representation sequencing agreed completely with identifications provided by whole-genome sequencing. Novel species were discovered within this set of bacterial isolates. A phylogenetic analysis and comparative genomic study of the three thermophilic bacterial genera, *Anoxybacillus*, *Geobacillus* and *Parageobacillus,* was performed to confirm the taxonomic placement of seven new genomes

of thermophilic bacteria. Substantial changes to the delimitation of the three genera have been made in recent years, and an integrated phylogenomic analysis was considered necessary to explore the phylogenetic relationships between these closely related genera, and provide correct placements for the newly sequenced genomes. A total of 113 complete genome assemblies from the RefSeq database, including *Anoxybacillus*, *Geobacillus* and *Parageobacillus*, were selected. Phylogenomic metrics were obtained, including calculation of Average Nucleotide Identity (ANI) and Average Amino acid Identity (AAI) and a maximum likelihood tree was constructed from alignment of a set of 662 orthologous core genes. The combined results from the core gene trees and ANI and AAI UPGMA dendrograms show that the genomes split into two main clades. Clade I contains all *Geobacillus*, all *Parageobacillus* and some species of *Anoxybacillus*, and Clade II, contains the majority of *Anoxybacillus* species. Clade I is further partitioned into three clades, consisting separately of *Geobacillus*, *Parageobacillus*, and a third clade which we suggest should be elevated to a new genus (*Quasigeobacillus gen. nov.*). In conclusion, complexity-reduced genotyping offers an accurate alternative to conventional methods for bacterial identification and strain typing and generates sequencing results without the need for previous sequence information for primer design. This allows for high-resolution sequence data to be produced for any bacteria without prior knowledge of taxonomic affinity. This technology fills a gap in currently available technologies, until such time as whole-genome sequencing is economically viable for routine application, and bioinformatic tools for such a purpose, are readily available for use.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Thanks to the Institute for Applied Ecology (IAE) for making my second half of the PhD a better experience. I would like to thank Arthur Georges, Ross Thompson, Susan Ward and Barbara Harris for supporting me during the administrative process in these 2.5 years of PhD. Without the IAE, I would have been deported by the University in 2017.

I would like to thank Prof. Arthur Georges for providing the guidance, support, knowledge, patience and resources to accomplish this research. Thanks for facilitating key elements for this project such as the approval for whole-genome sequencing, the acquisition of a genomic DNA certified reference of *E. coli* and facilitating access to books and scientific papers. It has been an enjoyable experience to be part of Arthur's research team. Knowledge, experience, curiosity and openness to new ideas makes Arthur an excellent supervisor.

I would like to acknowledge Consejo Nacional de Ciencia y Tecnología (CONACYT) for providing a scholarship "Becas CONACYT al extranjero 2015" to pursue graduate studies.

I would like to acknowledge Peter Wenzl for providing me the opportunity of receiving a training in Australia. Thanks to Peter and his team for reading my curriculum and deciding that I was a good candidate for that position. That decision changed my whole life.

I would like to thank Ling Xia for supporting me in my early beginnings of the PhD. When I had to attend meetings, workshops or do any lab work to look after my samples, she will allow me to attend without putting a complaint. Also, I would like to thank Cina for being friendly with me at the lab and Damian for proving guidance through some lab experiments and providing feedback of my PowerPoint slides prior my final seminar. Also, Andrzej Kilian for

paying the first semester of my PhD and allowing me to perform the experimental part of my PhD in DArT laboratories.

I would like to thank Dr M.A.(Rien) Habermehl for sharing his expertise and advice during the selection of sampling sites and indicating the best season for collecting samples in the Great Artesian Basin. I would like to thank Dinesh Shrestha for collecting and sending back to us, mud and water samples from Birdsville and Stoney crossing artesian water bores. Jason Carling and myself would like to thank the station managers for allowing us to collect samples from their artesian water bores. I also would like to thank Duncan's Plumbing service for their valuable help in collecting samples.

I would like to thank Karina Kennedy, Susan Bradbury and at the Canberra Public Hospital for providing me some of their samples and their identification results for my research.

I would like to thank Alan Hinge for the orientation and guidance provided in the early years of this journey. Alan's advices changed a lot my perspective of Australian culture. Also, I wold like to thank Joelle Vandermensbrugghe for advising me. I would like to acknowledge Margaret Chua from the UC medical and counselling for listening to me and doing some calls before the University tried to deport me for the 4th time. Margaret taught me the mindfulness meditation exercises, and interestingly I became more productive when I was able to afford an ergonomic chair, a desktop computer with enough capacity to run analysis and practiced meditation regularly. I would like to thank the Staff of the UC fit gym as they always welcomed me with a big smile. I would like to express my appreciation to the University of Canberra Council. Being member of the council during 2018 gave me a valuable experience that taught me the procedures of decision-making and governability of educative institutions. This was my first experience as a board member in a large institution.

While studying an MBA in Perú, I used to think that living there was going to be the hardest part of my life as student, but I was wrong. Nothing prepared me for what was about to come while being Australia. In these years as a PhD student, I met a broad range of people in both extremes, from the good, friendly and smart people, to the hostile and nasty people, and all classifications in between. All of them taught me something that I can summarize in one phrase that I will remember each time I have to take a decision, "Trust your guts".

I would like to acknowledge my Facebook friends for reading my frustration stories and writing nice comments to support me during the PhD. Also acknowledge the lady from Rosie's Chicken for giving me some free pieces of chicken when I didn't have money and I was hungry, and also choosing the "not-hard and dry" rice from the food tray. Also acknowledge Jeff from the Canberra City Gymnastics Club for giving me discounts during school term classes, and Rob for teaching me how to do a proper handstand. I would like to thank Josué García L for providing me with a good training program and a diet, as I am really excited that I will be on stage in 2 weeks.

I would like to acknowledge all my pets, current ones and the one that are not with us anymore, also the pets that are not mine but they are cute like Capitan Jack Jack and The longuies. The list includes Abu, Colachina, Prisca, Pinky, Nala, Percy, Tohui, Negra, Perry, Cameron, Ginger, Tiger, Rossy Bubbles, Chingis, Prietita, la werita, el gatito Prieto, Purruna, las tres rosetitas, Dollar, Boldo, Feo, Yeyo, Yeya, la comadre, Nutella, Miau, the compost worms in the garden, my kéfir grains, the magpie with one broken leg, the magpies around the block, Pet and Tortuga Rodríguez.

Thanks to my sisters for greeting me or calling me from time to time to check if everything was "ok" on the other side of the world (Australia). Maybe soon, we all three sisters will be called "Doctor".

One of the most important acknowledgements is for my parents for supporting my decision to study a PhD and giving me a most valuable asset, which is my education. Everything I am is thanks to them. Thanks to my mom for always giving me her honest opinion and thanks to my father for nurturing my analytical skills. At the beginning of my PhD in early 2015 my father was diagnosed with leukemia. These years have been challenging, because I have not been able to help from a distance. The fear of checking emails each morning asking if he improved after each chemotherapy, the nightmare of finding distributors of his medication from overseas, realizing that each flask of medication is almost $1000 dollars and he needed two per month and finding suitable plasma donors, is something I would not wish to any family. A few days before completing this document, he has been told that he may need more treatments. I may trust that medical advances will provide a cure and he will recover.

Finally, the other most important acknowledgement is for Jason, who has been there for me, before I decided to start a PhD and nothing of this would be happening without him. Jason has seen the best and the worst of me while being in Australia. From the joy of winning toys in the claw machine, to myself crying in the toilets of building 7 because I couldn't handle all my issues, from being knocked off the by a car, to changing my supervisory panel, to the treatment of my father with leukemia, to the rehabilitation after my accident, to start a PhD project from scratch in the second year, to the threats of deportation made the University. Thanks to Jason's patience and resilience to deal with my break downs each time I was thinking I was not going to finish this research on time. Jason has taken important roles in my life; his main role is as a much-loved partner. Then all his non-official roles that he has taken each time I ask: "Jason, can you help me with this?" and he ends up being my personal driver, English teacher, plumber, psychologist, gardener, chef, credit facilitator, physiotherapist, hairdresser, personal shopper, image consultant, life coach, photographer, mechanic, repairman of laser hair removal machines, cat catcher, story teller, musician, official cockroach and spider relocator, gymnastics

coach, trampoline coach, tourist guide and ranger in Kakadu, Rotorua and the Birdsville track, and over all, my unlimited source of hugs, kisses, stories and jokes that make me smile. I still remember the day I met you, 16 December 2013 in the kitchen of the CSIRO building at midday. I'm really excited for our wedding that will be next month.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. It is however, an industry supported PhD, having being provided the opportunity and logistic support of Diversity Arrays Technology Pty Ltd. The company provided sequencing within their facility at cost. This PhD was supported by a postgraduate research award from Consejo Nacional de Ciencia y Tecnología (CONACYT).

RESEARCH OUTPUTS DERIVED FROM THIS THESIS

1. Talamantes-Becerra, B., Carling, J., Kennedy K., Gahan, M., Georges, A. **2019**. Identification of bacterial isolates from a public hospital in Australia using complexity-reduced genotyping. Journal of Microbiological Methods Volume 160, Pages 11-19 DOI: https://doi.org/10.1016/j.mimet.2019.03.016
2. Talamantes-Becerra, B., Carling, J., Kennedy K., Gahan, M., Georges, A. **2019**. Short-read fastA files dataset from complexity-reduced genotyping by sequencing data of bacterial isolates from a public hospital in Australia. Data in Brief. Volume 25, 104273 DOI: https://doi.org/10.1016/j.dib.2019.104273
3. Talamantes-Becerra, B., Carling, J., Kilian, A., Georges, A. **2019**. Discovery of thermophilic *Bacillales* using reduced-representation genotyping for identification. BMC Microbiology ("*Submitted*").
4. Talamantes-Becerra, B., Carling, J., Blom, J., Georges, A. **2019**. Phylogenetic study of thermophilic genera *Anoxybacillus*, *Geobacillus* and *Parageobacillus* ("*In preparation for submission*").

CONFERENCE PRESENTATIONS

1. POSTER: Talamantes-Becerra, B., McNevin, D., Kilian, A., Carling, J., Georges, A. **2016**. DArTseq as a method for characterization and identification of bacterial isolates. Australian Society for Microbiology Annual Scientific Meeting, *Perth, Australia.*
2. SPEAKER: Talamantes-Becerra, B., Carling, J., Kilian, A., Gahan, M., McNevin, D., Georges, A. **2018**. DArTseq genotyping for bacterial identification: testing of complexity reduced methods. Genetics Society of Australasia, *Canberra, Australia*.

OTHER PUBLICATIONS

1. Talamantes-Becerra, B., González-Enríquez, R. **2019**. Mexican Journal of Biotechnology. Biosorption of lead with immobilised yeast in Sonora river water samples ("*In preparation for submission*").
2. Talamantes-Becerra, B and Nayyar, S. **2019**. Tiny Creatures Beneath Your Feet. Library for All Ltd, ISBN: 9781925960235

LIST OF FIGURES

LIST OF TABLES