DIVERSITY OF QUORUM SENSING PHEROTYPES AMONGST ECOTYPES OF PLANT-ASSOCIATED *BACILLUS SUBTILIS SENSU LATO* ISOLATES

by

HEATHER R. TREDGOLD

Submitted in fulfilment of the academic requirements of

Doctor of Philosophy

in Microbiology

School of Life Sciences
College of Agriculture, Engineering and Science
University of KwaZulu-Natal
Pietermaritzburg
South Africa

April 2021

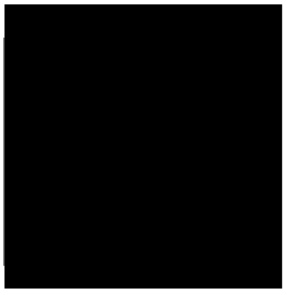
PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg Campus, South Africa. The research was financially supported by the National Research Foundation (Grant UID: 89632).

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

Signed: Dr. R. Hewer (Supervisor)

Date: April 2021



Signed: Dr. C. H. Hunter (Co-supervisor)

Date: April 2021

DECLARATION 1: PLAGIARISM

I, Heather Rayne Tredgold, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: Heather R. Tredgold (Candidate)

Date: April 2021

DECLARATION 2: PUBLICATIONS

Chapter 2

Tredgold, H. R., Hunter, C. H., Hewer, H. (2016) Evidence for a gene sequence basis for ecotype and pherotype determination amongst selected *Bacillus* spp.. Poster presentation to the 25th South African Society of Biochemistry and Molecular Biology Congress, 10–14 July 2016, East London, South Africa. Poster presented by H. R. Tredgold.

Tredgold, H. R., Hunter, C. H., Hewer, H. (2017) Evidence for a gene sequence basis for ecotype and pherotype determination amongst selected *Bacillus* spp.. Poster presentation to the University of KwaZulu-Natal School of Life Sciences Annual Post Graduate Research Day, 23 May 2017, Durban, South Africa. Poster presented by H. R. Tredgold.

Chapter 3

Tredgold, H. R., Hewer, H. , Hunter, C. H. (2020) Establishing ecotype variation amongst plant-associated isolates of the *Bacillus subtilis sensu lato* group demonstrating biocontrol potential. *Systematic and Applied Microbiology*. Submitted manuscript.

Chapter 4

Tredgold, H. R., Hewer, H. , Hunter, C. H. (2017) Pherotyping of plant-associated *Bacillus subtilis* and *B. amyloliquefaciens* isolates using PCR-RFLP analysis of *com*QXPA gene cluster. Poster presentation to the 6th American Society for Microbiology Conference on Cell–Cell Communication in Bacteria, 16–19 October 2017, Athens, Georgia, U. S. A. Poster presented by H. R. Tredgold.

Tredgold, H. R., Hewer, H. , Hunter, C. H. (2017) Pherotyping of plant-associated *Bacillus subtilis* and *B. amyloliquefaciens* isolates using PCR-RFLP analysis of *com*QXPA gene cluster. Poster presentation to the University of KwaZulu-Natal College of Agriculture, Engineering and Science Postgraduate Research and Innovation Day, 26 October 2017, Durban, South Africa. Poster presented by H. R. Tredgold.

Signed: Heather R. Tredgold (Candidate)

Date: April 2021

ABSTRACT

Ecologically-adapted populations, or ecotypes, are species forms that are functionally adapted to particular habitat niches. Bacterial ecotypes are challenging the significance and implications of ecological adaptation in terms of prokaryote taxonomy, community ecology, and biocontrol applications. Plant-associated members of the *Bacillus subtilis sensu lato* group—in particular *B. subtilis* and *B. velezensis*—are ecologically specialised to perform numerous functions that are beneficial to plant and soil health. Underpinning these beneficial activities is the ability to colonise plant surfaces by biofilm formation. Biofilms are a result of co-ordinated social behaviour amongst microorganisms. In the *B. subtilis sensu lato* this sociality is governed by the ComQXPA quorum-sensing cassette, which uses the ComX pheromone for intercellular communication leading social behaviours like biofilm formation. This peptide pheromone contains a post-translational modification on a conserved tryptophan residue. This modification is highly variable between populations, resulting in discrete ComX variants which form communication groupings known as pherotypes. The limitation of communication to within a pherotype may constitute a form of ecological adaptation designed to protect the products of co-operative behaviour and restrict their benefit to the producing population.

The present study aimed to explore ecotypes and pherotypes amongst a subset of plant-associated *B. subtilis* and *B. velezensis* isolates. These isolates originated from phylloplane and rhizosphere samples from seven crop species grown across the KwaZulu-Natal province, South Africa, and had demonstrated biocontrol potential in previous studies. An exploratory study set out to apply in silico approaches to determine gene-sequence-based variation amongst representative strains of the *B. subtilis sensu lato* available in the GenBank database. Nine housekeeping gene targets (viz., 16S rRNA, *cheA*, *dnaJ*, *groEL*, *gyrA*, *gyrB*, *polC*, *purH*, and *rpoB*) were evaluated for suitability to resolve clustering of closely-related *B. subtilis sensu lato*. Four of these gene sequences (viz., 16S rRNA, *dnaJ*, *gyrA*, and *rpoB*) were identified as candidates for a Multilocus Sequence Analysis (MLSA) scheme to distinguish between members of the *B. subtilis sensu lato* group. Putative pherotype variation amongst these reference strains was explored in silico using *comQXP* gene sequence data. The suitability of a *comQXP* PCR-RFLP protocol with potential for rapid pherotyping amongst *B. subtilis sensu lato* was evaluated in silico using simulated *comQXP* amplicons. This necessitated the design

of a PCR primer set targeting the quorum-sensing gene region of *B. velezensis*. Four restriction enzyme candidates namely, *Bts*CI, *Fnu*4HI, *Cac*8I, and *Hpy*166II, were identified for further study.

Ecotyping amongst the *B. subtilis* and *B. velezensis* isolates was carried out using a four-gene (viz., 16S rRNA, *dnaJ*, *gyrA*, and *rpoB*) MLSA. This concatenated sequence dataset was applied to ecotype simulation (ES) analysis to corroborate putative ecotype sub-clusters in the MLSA phylogeny. Two DNA fingerprinting approaches—Repetitive Element Palindromic PCR (Rep-PCR) and Random Amplified Polymorphic DNA PCR (RAPD-PCR)—were also evaluated for their potential to identify putative ecotypes within the isolate subset. This investigation of pherotypes amongst the putative ecotype groupings examined isolate *com*Q sequence data as well as the *comQXP* PCR-RFLP, and also applied a *srfA*-LacZ reporter gene assay to examine isolate stimulation of seven known pherotype tester strains (viz., 168, RO-B-2, RO-C-2, RO-E-2, RO-FF-1, RO-H-1, and RS-D-2).

The MLSA of isolate gene sequences presented distinct sequence clusters suggestive of ecotype populations amongst the two *Bacillus* species which were corroborated by ES analysis. The MLSA and ES determined two putative ecotypes within the *B. subtilis* isolates, and six within the *B. velezensis* isolates. Ecotype groupings were found to contain isolates from different crop species and locations, and four *B. velezensis* ecotypes were distinct from included *B. subtilis sensu lato* reference strains. PCR fingerprinting identified strain-level variances amongst the isolates, and were able to differentiate plant-associated B. *velezensis* from closely-related *B. amyloliquefaciens*, but did not define groupings entirely consistent with the ES and MLSA phylotypes. The MLSA, ES, and PCR fingerprinting delineated a group of five isolates (viz., B81, CT-R67, bnd136, bng221, and sqo271) to be grouped with reference strain *Bacillus* sp. JS in the MLSA. This grouping is of interest as *Bacillus* sp. JS is a strain demonstrating biocontrol capability and formed a clade distinct to *B. subtilis sensu stricto* in gene sequence phylogenies. This *Bacillus* sp. JS grouping was further confirmed by ES to constitute a single ecotype, and both Rep-PCR and RAPD-PCR OPG-16 distinguished this grouping based on fingerprint profiles.

Investigation of reference strain variation in the *comQ* gene sequence determined significant levels of sequence variation amongst reference strains evaluated in silico, and showed divergence in some strains from known pherotype counterparts. Furthermore, sequence clusters were resolved in *B. velezensis* that suggested unique pherotype variants amongst reference strains, and showed *comQ* similarity amongst the five isolates related to *Bacillus* sp. JS. This trend was observed for the *comQ* phylogeny applying isolate and pherotype sequences, which resolved two distinct clusters each amongst the *B. velezensis* and *B. subtilis* isolates, with only two isolates (viz., bnd134 and bnd156) grouping with known pherotypes. The *B. velezensis* clades demonstrated significant *comQ* sequence variance from the pherotype reference strains based on clustering distances in the phylogeny.

Of the four enzyme candidates identified for potential use in a *comQXP* PCR-RFLP protocol, two were found to be applicable: *Bts*CI resolved the grouping of the *Bacillus* sp. JS clade, while *Fnu*4HI was found to distinguish between more closely-related isolates in *B. velezensis*. The profile groupings for both of these enzymes demonstrated most of the isolates to be distinct from known reference pherotype strains, and the technique proved able to corroborate groupings in the *comQ* phylogeny. The pherotyping *srfA*-LacZ reporter gene assay failed to support the pherotype groupings defined by the *comQ* phylogeny, but demonstrated that the environmental isolates were capable of eliciting substantial responses in the pherotype tester strains, particularly amongst the *B. velezensis* isolates. Eight isolates did not elicit significant responses in any of the tester strains, while most matched to multiple pherotypes on the basis of tester strain response to conditioned medium from isolates.

These findings demonstrated that putative ecotype variation was present amongst the *B. subtilis* and *B. velezensis* isolates that were resolvable by MLSA, and that these ecotype groups comprised isolates from different crop types and location sites. Some of these ecotypes bore only distant relation to *B. subtilis sensu lato* reference strain counterparts. The investigation of PCR fingerprinting methods for ecotyping purposes found that the primer sets applied were not able to consistently corroborate putative ecotype groupings defined by MLSA and ES. Pherotyping investigations demonstrated that there was evidence of gene-sequence-based pherotype variation amongst the isolates within these putative ecotype groupings. The

pherotype clades resolved with the *comQ* phylogeny and with *comQXP* PCR-RFLP were not corroborated by *srfA*-LacZ reporter gene pherotyping assays, but the data suggested that assay optimisation to suit environmental strains is warranted in future studies.

This study presents the first instance of ecotyping and pherotyping amongst plant-beneficial *B. velezensis*. The *B. velezensis* isolates were largely distinct in ecotype and pherotype from the plant-associated model strain FZB42, which further promotes that these isolates demonstrate localised adaptation. The *Bacillus* sp. JS-related isolate grouping is of interest as these isolates comprised a single putative ecotype resolvable by MLSA, ES, and Rep-PCR and RAPD-PCR OPG-16. This grouping was distantly-related to *B. subtilis sensu stricto*, and this is interesting from a taxonomic standpoint, as is the fact that that *Bacillus* sp. JS is reported to have plant-beneficial capabilities. The existence of uniquely South African plant-associated *B. subtilis sensu lato* ecotypes is valuable in agricultural management approaches targeting beneficial microbes, as these organisms likely possess adaptations allowing them to compete with extant bacterial communities and thrive within the South African agricultural climate.

ACKNOWLEDGMENTS

This research was generously funded by the National Research Foundation. I would also like to extend my gratitude to Dr. David Dubnau (Public Health Research Institute, New Jersey Medical School, Rutgers University, U. S. A.) for his generous gift of pherotype reference strains for use in this research.

I would like to offer special thanks to my project supervisors Dr. Charles Hunter and Dr. Raymond Hewer for their advice and abundant encouragement during the course of this study.

The tireless behind-the-scenes work conducted by the dedicated administrative and support staff of the School of Life Sciences is most gratefully acknowledged. Thank you to Celeste Clark, Di Fowlds, Megan Brunkhorst, and Pat Joubert for the professional and personal support they made freely available to me.

Many academic staff offered their advice and encouragement, and I would like to especially thank Prof. Dean Goldring for encouraging my writing endeavours and including me in his writing group and workshops, and Prof. Theresa Coetzer for being so accommodating and welcoming of my presence in her lab.

I wish to thank my postgraduate colleagues within the School of Life Sciences for their support and camaraderie over the years. I am most thankful to my lab-families in Lab 223 and Lab 46. My cheerleader-in-chief, Dr. Lauren Eyssen, deserves extra-special mention.

My family, both local and far-flung, have offered their unwavering support during my postgraduate endeavours, for which I am forever grateful. Mom, thank you for your guidance and love during this venture and always.

My final thanks go to my partner Matthew for taking this journey with me, and for his many roles as fellow scientist, advisor, sounding-board, sympathetic ear, and motivator. I truly appreciate you.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

γ-PGA	Poly-gamma-glutamic acid
AAI	Average amino acid identity
AHLs	Acyl homoserine lactones
AIPs	Autoinducer peptides
ANI	Average nucleotide identity
ATCC	American Type Culture Collection
BGSC	Bacillus Genetic Stock Center
BLAST	Basic local alignment search tool
CSF	Competence and sporulation stimulating factor
dDDH	Digital DNA-DNA hybridisation
DDH	DNA-DNA hybridisation
DNA	Deoxyribonucleic acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EDTA	Ethylenediaminetetraacetic acid
ECM	Extracellular matrix
EPS	Extracellular polymeric substances
ES	Ecotype simulation
LB	Luria-Bertani broth
MEGA	Molecular evolutionary genetics analysis
MLSA	Multilocus sequence analysis
MLST	Multilocus sequence typing
MUSCLE	Multiple sequence comparison by log-expectation
MWM	Molecular weight marker
NCBI	National Center for Biotechnology Information
OD	Optical density
ONPG	*ortho*-Nitrophenyl-β-galactoside
PCR	Polymerase chain reaction
Rap	Response-regulator aspartyl-phosphate phosphatases
RAPD	Randomly amplified polymorphic DNA
Rep	Repetitive extragenic palindromic

RFLP	Restriction fragment length polymorphisms
rRNA	Ribosomal RNA
SBA	Sodium boric acid buffer
TBE	Tris-borate-ethylenediaminetetraacetic acid buffer
TETRA	Tetranucleotide signatures
TSA	Tryptic soy agar
TSB	Tryptic soy broth
UPGMA	Unweighted pair group method with arithmetic mean

INTRODUCTION

Microbial ecology studies in recent decades have expanded our understanding of community dynamics and species evolution, and one aspect of this debate is the existence and relevance of the ecological species concept amongst bacterial species (Koeppel *et al*., 2008). Ecological gradients in molecular diversity and subpopulation level clustering within housekeeping gene phylogenies have been interpreted across several studies to represent "ecologically adapted species populations" or ecotypes (Cohan, 2005, Ward *et al.*, 2007, Koeppel *et al.*, 2008). Ecotypes are important as they account for variations in functional diversity which allow bacterial populations to expand into and colonise their chosen niche (Lan and Reeves, 2000). The bacterial ecotype concept then reflects diversity within prokaryote species which has evolved as a result of ecological selection pressure (Ward, 2006).

Bacillus species are well known for their habitat diversity and wide range of metabolic capabilities (Mandić-Mulec *et al.*, 2015). Despite the low level of discrimination offered by the 16S rRNA gene sequence, *Bacillus* spp. exhibit large degrees of heterogeneity in other functional genes (Mandić-Mulec *et al*., 2015). Such genetic variations have resulted in identified groupings of related taxa amongst *Bacillus* spp., one example being the *B. subtilis sensu lato* group (Alcaraz *et al*., 2010, Bhandari *et al*., 2013, Fajardo-Cavazos *et al*., 2014). This particular group comprises many species of environmental importance, and has been significantly expanded and revised over the last decade (Fajardo-Cavazos *et al*., 2014). Such *sensu lato* groupings account for the high degree of genomic variance between closely-related species, and this also reflects a degree of divergence in ecological terms. The sequencing of additional core genes and application of approaches such as Multilocus Sequence Analysis (MLSA) and genomic fingerprinting have presented evidence for diversification into ecologically distinct groupings (i.e. ecotypes) within members of the *B. subtilis* group of related taxa (Istock *et al*., 2001, Reva *et al*., 2004, Connor *et al*., 2010, Mandić-Mulec *et al*., 2015). The recent taxonomic debate surrounding plant-associated *B. velezensis* in particular underlines the impacts of ecological variation within the members of this grouping (Dunlap *et al*., 2015, Dunlap *et al*., 2016).

Several members of the *B. subtilis sensu lato* grouping—in particular *B. subtilis* and *B. velezensis*—are highly regarded for existence in association with plants and roles in plant health maintenance (Govindasamy *et al.*, 2010, Borriss, 2011, Fan *et al.*, 2018). Like many beneficial bacterial species, *Bacillus* spp. existing on plant surfaces form cooperative biofilms which allow them to colonise and persist within their niche (Rudrappa *et al.*, 2008, Altaf *et al.*, 2017). Biofilm formation and maintenance pathways in *Bacillus* are highly complex and overlap many other pathway systems, but are underpinned by the quorum sensing pheromone ComX (Solomon *et al.*, 1995, Solomon *et al.*, 1996, Solomon and Grossman, 1996, Hamoen *et al.*, 2003). The ComX autoinducer peptide is synthesised by a signalling cassette encoded in the *com*QXPA gene loci (Dogsa *et al.*, 2014). ComX pheromones are ecologically interesting in that they are highly distinct between populations of the same species, as a result of variable posttranslational modifications (Ansaldi *et al.*, 2002, Ansaldi and Dubnau, 2004). These modifications involve the addition of an isoprenoid group to a conserved tryptophan residue, with the nature of the isoprenylation and the length of the linked side chain being hypervariable (Okada, 2011). ComX variation then represents a form of social grouping intended to protect "public goods" (i.e. the products of social behaviour which benefit the producing population) and ensure cooperation amongst the discrete population (Smith and Schuster, 2019). To date several distinct pherotype groupings have been identified within the *B. subtilis sensu lato* (Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002, Okada *et al.*, 2015, Esmaeilishirazifard *et al.*, 2017).

The *com*QXPA quorum-sensing system represents a probable niche or speciation mechanism on the basis of the structural variety in quorum sensing pheromones and their cognate receptors, and with the co-evolution of the genes encoding the pheromone (ComX), processing enzyme (ComQ), and binding portions of the histidine kinase receptor (ComP) (Tortosa and Dubnau, 1999). The majority of studies of ComX variation and mechanisms have been restricted to a subset of well-known pherotype reference strains (Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002, Bacon Schneider *et al.*, 2002, Ansaldi and Dubnau, 2004), and expansion into environmental studies have been carried out on soil-derived B. *subtilis sensu lato* isolates (Stefanic and Mandić-Mulec 2009) and those originating from the tomato rhizoplane (Oslizlo *et al.*, 2015). To date, no studies have been carried out to evaluate the link between ecotype and pherotype variation in plant-associated *B. subtilis sensu lato* from various localities and plant origins. Furthermore, no research has been conducted on ecotype or ComX variation amongst plant-

associated *B. velezensis*, despite their renown as beneficial plant-associated bacteria (Rabee *et al*., 2019, Reva *et al*., 2019).

The existence of social grouping variation in the form of distinct ComX pheromones and resultant highly-specific quorum-sensing responses affords the grounds to propose a link between the population-specificity of ComX pheromones and ecological diversification in plant-associated *B. subtilis sensu lato* (Stefanic *et al*., 2012, Mandić-Mulec *et al*., 2015). This study aims to define ecotypes amongst a subset of plant-associated *B. subtilis* and *B. velezensis* isolates, and to investigate pherotype variation within the ecotype groupings of this subset. The research objectives of this study may be summarised as follows: -

1) Use in silico approaches to evaluate variation in quorum-sensing gene loci of selected *B. subtilis sensu lato* reference strains, and explore housekeeping gene candidates to resolve differences between closely-related members of the *B. subtilis sensu lato*.
2) Evaluate ecotypes amongst *B. subtilis* and *B. velezensis* isolates demonstrating biocontrol potential.
3) Determine pherotype diversity amongst the putative ecotype groupings of *B. subtilis* and *B. velezensis* isolates.

The dissertation has been divided into five chapters. Chapter one presents a literature review exploring the sociality of *B. subtilis* as biofilms and the role of the *comQXP* quorum-sensing cassette, the phenomenon ecological adaptation in bacteria and its presence in *B. subtilis*, pherotype variation in ComX, and the relationship between ecotypes and pherotypes in the *B. subtilis sensu lato*. Chapter two presents the findings of the in-silico aspects of this study, and chapters three and four present the respective investigations into ecotypes and pherotypes amongst the plant-associated *B. subtilis sensu lato* isolates which comprise the focus of this research. In the final fifth chapter, a general overview of the research and examination of significant findings is presented.

References

Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L. & Olmedo, G. (2010) Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* **11**: 332, 1–17. 10.1186/1471-2164-11-332

Altaf, M. M., A., I., Khan , M. S. A. & Grohmann, E. (2017) *Bacillus* biofilms and their role in plant health. *Biofilms in Plant and Soil Health* (Ahmad, I. & Husain, F. M., eds.), pp. 55–67. https://doi.org/10.1002/9781119246329.ch4

Ansaldi, M. & Dubnau, D. (2004) Diversifying selection at the *Bacillus* quorum-sensing locus and determinants of modification specificity during synthesis of the ComX pheromone. *Journal of Bacteriology* **186**: 1, 15–21. 10.1128/jb.186.1.15-21.2004

Ansaldi, M., Marolt, D., Stebe, T., Mandić-Mulec, I. & Dubnau, D. (2002) Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Molecular Microbiology* **44**: 6, 1561–1573. https://doi.org/10.1046/j.1365-2958.2002.02977.x

Bacon Schneider, K., Palmer, T. M. & Grossman, A. D. (2002) Characterization of *comQ* and *comX*, two genes required for production of ComX pheromone in *Bacillus subtilis*. *Journal of Bacteriology* **184**: 2, 410–419. 10.1128/JB.184.2.410-419.2002

Bhandari, V., Ahmod, N. Z., Shah, H. N. & Gupta, R. S. (2013) Molecular signatures for *Bacillus* species: Demarcation of the *Bacillus subtilis* and *Bacillus cereu*s clades in molecular terms and proposal to limit the placement of new species into the genus Bacillus. *International Journal of Systematic and Evolutionary Microbiology* **63**: 7, 2712–2726. 10.1099/ijs.0.048488-0

Borriss, R. (2011) Use of plant-associated *Bacillus* strains as biofertilizers and biocontrol agents in agriculture. *Bacteria in Agrobiology: Plant Growth Responses* (Maheshwari, D. K., ed.) pp. 41–76. Springer-Verlag, Berlin, Germany. 10.1007/978-3-642-20332-9_3

Cohan, F. M. (2005) Periodic selection and ecological diversity in bacteria. *Selective Sweep* (Nurminsky, D., ed.) pp. 78–93. Springer US, Boston, MA. 10.1007/0-387-27651-3_7

Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koeppel, A. F., Burger, A., Cole, S. G., Perry, E. B., Krizanc, D., Field, N. C., Slaton, M. & Cohan, F. M. (2010) Ecology of speciation in the genus *Bacillus*. *Applied and Environmental Microbiology* **76**: 5, 1349–1358. 10.1128/aem.01988-09

Dogsa, I., Oslizlo, A., Stefanic, P. & Mandić-Mulec, I. (2014) Social interactions and biofilm formation in *Bacillus subtilis*. *Food Technology and Biotechnology* **52**: 2, 149–157. 10.1128/JB.01290-08

Dunlap, C. A., Kim, S.-J., Kwon, S.-W. & Rooney, A. P. (2015) Phylogenomic analysis shows that *Bacillus amyloliquefaciens* subsp. *plantarum* is a later heterotypic synonym of *Bacillus methylotrophicus*. *International Journal of Systematic and Evolutionary Microbiology* **65**: 7, 2104–2109. doi:10.1099/ijs.0.000226

Dunlap, C. A., Kim, S.-J., Kwon, S.-W. & Rooney, A. P. (2016) *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and *'Bacillus oryzicola'* are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. *International Journal of Systematic and Evolutionary Microbiology* **66**: 3, 1212–1217. doi:10.1099/ijsem.0.000858

Esmaeilishirazifard, E., De Vizio, D., Moschos, S. A. & Keshavarz, T. (2017) Genomic and molecular characterization of a novel quorum sensing molecule in *Bacillus licheniformis*. *AMB Express* **7**: 78, 1–13. 10.1186/s13568-017-0381-6

Fajardo-Cavazos, P., Maughan, H. & Nicholson, W. L. (2014) Evolution in the *Bacillaceae*. *Microbiology Spectrum* **2**: 5, 1–32. 10.1128/microbiolspec.TBS-0020-2014

Fan, B., Wang, C., Song, X., Ding, X., Wu, L., Wu, H., Gao, X. & Borriss, R. (2018) *Bacillus velezensis* FZB42 in 2018: The Gram-positive model strain for plant growth promotion and biocontrol. *Frontiers in Microbiology* **9**: 2491. 10.3389/fmicb.2018.02491

Govindasamy, V., Senthilkumar, M., Magheshwaran, V., Kumar, U., Bose, P., Sharma, V. & Annapurna, K. (2010) *Bacillus* and *Paenibacillus* spp.: Potential PGPR for sustainable agriculture. *Plant Growth and Health Promoting Bacteria* (Maheshwari, D. K., ed.) pp. 333–364. Springer-Verlag, Berlin, Germany. https://doi.org/10.1007/978-3-642-13612-2

Hamoen, L. W., Venema, G. & Kuipers, O. P. (2003) Controlling competence in *Bacillus subtilis*: Shared use of regulators. *Microbiology* **149**: 9–17. 10.1099/mic.0.26003-0

Istock, C. A., Ferguson, N., Istock, N. L. & Duncan, K. E. (2001) Geographical diversity of genomic lineages in *Bacillus subtilis* (Ehrenberg) Cohn *sensu lato*. *Organisms Diversity & Evolution* **1**: 3, 179–191. https://doi.org/10.1078/1439-6092-00017

Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., Rooney, A. P., Brambilla, E., Connor, N., Ratcliff, R. M., Nevo, E. & Cohan, F. M. (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 7, 2504–2509. 10.1073/pnas.0712205105

Lan, R. & Reeves, P. R. (2000) Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends in Microbiology* **8**: 9, 396–401. 10.1016/S0966-842X(00)01791-1

Mandić-Mulec, I., Stefanic, P. & van Elsas, J. D. (2015) Ecology of *Bacillaceae*. *Microbiology Spectrum* **3**: 1, 1–23. 10.1128/microbiolspec.TBS-0017-2013

Okada, M. (2011) Post-translational isoprenylation of tryptophan. *Bioscience, Biotechnology, and Biochemistry* **75**: 8, 1413–1417. 10.1271/bbb.110087

Okada, M., Nakamura, Y., Hayashi, S., Ozaki, K. & Usami, S. (2015) Chemical structure and biological activity of a quorum sensing pheromone from *Bacillus subtilis* subsp. *natto*. *Bioorganic & Medicinal Chemistry Letters* **25**: 19, 4293–4296. https://doi.org/10.1016/j.bmcl.2015.07.083

Oslizlo, A., Stefanic, P., Vatovec, S., Beigot Glaser, S., Rupnik, M. & Mandić-Mulec, I. (2015) Exploring ComQXPA quorum-sensing diversity and biocontrol potential of *Bacillus* spp. isolates from tomato rhizoplane. *Microbial Biotechnology* **8**: 3, 527–540. 10.1111/1751-7915.12258

Rabbee, M. F., Ali, M. S., Choi, J., Hwang, B. S., Jeong, S. C. & Baek, K.-H. (2019) *Bacillus velezensis*: A valuable member of bioactive molecules within plant microbiomes. *Molecules* **24**: 6, 1046. 10.3390/molecules24061046

Reva, O. N., Dixelius, C., Meijer, J. & Priest, F. G. (2004) Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *FEMS Microbiology Ecology* **48**: 249–259. 10.1016/j.femsec.2004.02.003

Reva, O. N., Swanevelder, D. Z. H., Mwita, L. A., Mwakilili, A. D., Muzondiwa, D., Joubert, M., Chan, W. Y., Lutz, S., Ahrens, C. H., Avdeeva, L. V., Kharkhota, M. A., Tibuhwa, D., Lyantagaye, S., Vater, J., Borriss, R. & Meijer, J. (2019) Genetic, epigenetic and phenotypic diversity of four *Bacillus velezensis* strains used for plant protection or as probiotics. *Frontiers in Microbiology* **10**: 2610, 1–25. 10.3389/fmicb.2019.02610

Rudrappa, T., Biedrzycki, M. L. & Bais, H. P. (2008) Causes and consequences of plant-associated biofilms. *FEMS Microbiology Ecology* **64**: 2, 153–166. 10.1111/j.1574-6941.2008.00465.x

Smith, P. & Schuster, M. (2019) Public goods and cheating in microbes. *Current Biology* **29**: 11, R442–R447. https://doi.org/10.1016/j.cub.2019.03.001

Solomon, J. M. & Grossman, A. D. (1996) Who's competent and when: Regulation of natural genetic competence in bacteria. *Trends in Genetics* **12**: 4, 150–155. https://doi.org/10.1016/0168-9525(96)10014-7

Solomon, J. M., Lazazzera, B. A. & Grossman, A. D. (1996) Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes & Development* **10**: 16, 2014–2024. 10.1101/gad.10.16.2014

Solomon, J. M., Magnuson, R., Srivastava, A. & Grossman, A. D. (1995) Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes & Development* **9**: 5, 547–558. 10.1101/gad.9.5.547

Stefanic, P. & Mandić-Mulec, I. (2009) Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. *Journal of Bacteriology* **191**: 6, 1756–1764. 10.1128/jb.01290-08

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F. M. & Mandić-Mulec, I. (2012) The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. *Environmental Microbiology* **14**: 6, 1378–1389. 10.1111/j.1462-2920.2012.02717.x

Tortosa, P. & Dubnau, D. (1999) Competence for transformation: A matter of taste. *Current Opinion in Microbiology* **2**: 6, 588–592. https://doi.org/10.1016/S1369-5274(99)00026-0

Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandić-Mulec, I. & Dubnau, D. (2001) Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. *Journal of Bacteriology* **183**: 2, 451–460. 10.1128/JB.183.2.451-460.2001

Ward, D. M. (2006) A macrobiological perspective on microbial species. Vol. 1 pp. 269–278. *Microbe Magazine*. American Society for Microbiology. 10.1128/microbe.1.269.1

Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kuhl, M. & Grossman, A. (2007) Genomics, environmental genomics and the issue of microbial species. *Heredity* **100**: 2, 207–219. 10.1038/sj.hdy.6801011

CHAPTER ONE

LITERATURE REVIEW

1.1. Introduction

In the midst of a rapidly expanding global human population and the turmoil of climate change, food security is at a higher priority than ever before. Food security is particularly relevant in countries with large proportions of low-income groups — such as South Africa — where many families rely on subsistence agriculture to meet their daily dietary needs in the face of ever-increasing food costs. It was estimated that in 2016 there were over 800 million people facing chronic malnutrition globally, and 22.7% of sub-Saharan Africans were undernourished (FAO, 2017). These statistics are compounded by deteriorating climate conditions and conflict in already-at-risk areas such as sub-Saharan Africa and parts of Asia (FAO, 2017). Crop productivity and food accessibility issues are exacerbated by the expense and environmental risks associated with the use of conventional chemical fertilisers and agricultural pesticides. Investigations into the potential of microbes in place of costly agrichemicals to maintain soil health and combat crop pests and diseases has been widely explored and yielded several commercialised *Bacillus*-based products (Cawoy *et al.*, 2011). Modern agricultural practices are increasingly advocating and encouraging the use of alternative approaches to crop management, such as utilising microbial species as biofertilisers and biopesticides (Compant *et al.*, 2005). Aerobic endospore-forming bacteria—in particular members of the family *Bacillaceae* and genus *Bacillus*—play important roles in plant health and disease antagonism, and several species are commonly found in soil and in association with plants (Borriss, 2011). Members of the *Bacillaceae* family have been the subject of numerous ecological and biocontrol studies owing to their ubiquity, competitiveness in the plant-associated habitat, beneficial plant-health-promotion activities, and ease of formulation (Govindasamy *et al.*, 2010, Borriss, 2011, Cawoy *et al.*, 2011).

Biofilm formation by bacteria residing in soil- and plant-associated environments contributes significantly to soil health, nutrient cycling, and plant growth (Angus and Hirsch, 2013). The

ubiquity and persistence of *Bacillus* spp. in the soil- and plant-associated environment has been attributed to their abilities to form biofilms (Yaryura *et al.*, 2008, Chen *et al.*, 2013). Plant-associated biofilms are underpinned by plant exudates which promote *Bacillus* spp. to form biofilms and ensure root colonisation by desirable plant-associated species (Bais *et al.*, 2004, Yaryura *et al.*, 2008, Beauregard *et al.*, 2013, Chen *et al.*, 2013, Reva *et al.*, 2019). *Bacillus* spp. quorum sensing molecules and the pathways of biofilm formation, competence development, and sporulation have been the focus of numerous studies (Solomon *et al.*, 1995, Solomon *et al.*, 1996, Tortosa *et al.*, 2001, Hamoen *et al.*, 2003b, Schultz *et al.*, 2009, Parashar *et al.*, 2013).

Competence behaviours, sporulation, biofilm formation, and surfactin production in *B. subtilis sensu lato* are governed by the peptide pheromone ComX (Ansaldi *et al.*, 2002). The array of post-translational modifications to the ComX quorum-sensing pheromone have resulted in highly variable ComX between populations, and this has defined communication-specific pherotype populations in *Bacillus subtilis* and related species (Okada *et al.*, 2005, Pottathil *et al.*, 2008, Stefanic *et al.*, 2012). Privatisation in signal molecules serves to prevent competitors gaining advantages from cooperative activities and is useful in the competitive and diverse plant environment (Ansaldi and Dubnau, 2004). Specialisation into pherotypes has been linked to ecological adaptation to a particular niche and lifestyle (Williams *et al.*, 2007). Several analyses of the genetic relationships between environmental *Bacillus* spp. have found evidence for ecologically distinct groupings (i.e. ecotypes) within members of the *B. subtilis* group of related taxa (Reva *et al.*, 2004, Mandić-Mulec *et al.*, 2015). A link between the ecotype and pherotype concepts in *Bacillus* has been proposed by researchers, since an ecotype grouping represents a specialised niche adaptation, and a pherotype is a social grouping diversified on the basis of ComX pheromone population-specificity and the need for discrete communication (Reva *et al.*, 2004, Stefanic *et al.*, 2012, Mandić-Mulec *et al.*, 2015).

To date, little research examining these two concepts in the context of plant-associated *B. subtilis sensu lato* has been reported, and in particular there are no studies investigating these in plant-associated *B. velezensis*. The present study aims to investigate the existence of ecotypes and pherotypes amongst a subset of plant-associated wild-type *B. velezensis* and *B.*

subtilis isolates. To this end, this literature review explores the sociality of bacteria within biofilms and, in particular, examines the competence and biofilm formation system controlled by the ComQXPA signalling system of *Bacillus* spp.. Evidence for bacterial ecotypes, how these fit within the traditional microbial species concept, and the relationship between ecotypes and pherotypes in *Bacillus* spp. will also be discussed.

1.2. Quorum sensing and bacterial biofilms

Bacterial biofilms can be both beneficial and detrimental to human activities, and are relevant in industry, agriculture, and medicine (Murthy and Venkatesan, 2009, Shi and Zhu, 2009, Velmourougane *et al.*, 2017). Microbes are able to form biofilms on biotic and abiotic surfaces, and bacteria persist as biofilms in most natural settings (Ramey *et al.*, 2004, Lemon *et al.*, 2008). The actual metabolic processes and genes involved in biofilm formation and maintenance vary between bacterial species (Lemon *et al.*, 2008). Biofilms are formed by populations of surface-associated microbes which adhere to each other by a self-produced matrix to form structured communities of differentiated cells (Lemon *et al.*, 2008, Vlamakis *et al.*, 2013). Survival as biofilms is considered to be an attribute most microorganisms, as this state provides the cells with the competitive advantage of group behaviours such as cooperation for nutrition, resistance to antimicrobial substances, and protection from the surrounding environment (Lazazzera and Grossman, 1998, Ramey *et al.*, 2004, Lemon *et al.*, 2008).

Bacteria form biofilms as a response to environmental stimuli and self-produced quorum sensing signals (Camilli and Bassler, 2006). Quorum sensing enables a cell-density-dependent process of communication within a bacterial population to occur, which in turn facilitates coordination of community behaviour (Joint *et al.*, 2007, Williams, 2007). Quorum sensing systems have been described in most bacterial species and control the expression and regulation of 5–25% of genes in a bacterial cell (Joint *et al.*, 2007). These systems have been linked to the regulation of a variety of behaviours including virulence, genetic competence, secondary metabolite production, motility, symbiosis, and nodulation (Sturme *et al.*, 2002, Joint *et al.*, 2007, Stefanic and Mandić-Mulec, 2009, Dogsa *et al.*, 2014). Induced behavioural changes allow the bacteria to synchronise responses throughout the population and facilitate cooperative

functioning analogous to a multicellular organism, in order to reap benefits ordinarily unattainable to cells living independently (Schauder and Bassler, 2001, Waters and Bassler, 2005). The benefits of cooperation in microbes are termed "public goods", which are extracellular products that contribute to community survival and population density, such as nutritive elements, small proteins, and metabolites (Smith and Schuster, 2019).

Switches in gene expression allow bacteria to behave differently in the biofilm than they would in the free-living state (Henke and Bassler, 2004). Biofilm-associated cells exhibit physiological differences from those living planktonically in dispersed culture because biofilm formation requires the bacterial cell to switch from a unicellular (planktonic) state to a sedentary "multicellular" state (Ramey *et al.*, 2004, Lemon *et al.*, 2008). When desirable biofilm-formation conditions arise, the cells which exist in a non-motile unicellular state will produce surface adhesins to increase adhesion to other cells and surfaces; motile cells lose flagellar motility and begin to secrete an extracellular matrix (ECM) (Lemon *et al.*, 2008). The cells in a biofilm adhere by a self-produced ECM comprised of extracellular polymeric substances (EPS) comprising exopolysaccharides, exogenous DNA, and proteins (Branda *et al.*, 2005). Biofilms may assume a range of configurations, from flat and featureless, to tightly clustered aggregates, or complex structures such as towers or streamers (Ramey *et al.*, 2004). Five general stages have been identified in the formation of a bacterial biofilm (Lemon *et al.*, 2008): Cells attach to a surface, become sessile, forming a monolayer; more cells aggregate to form a multi-layered microcolony; cells begin to produce an extracellular matrix; as the biofilm matures, a three-dimensional architecture develops.

Bacterial quorum sensing uses both intra- and extra-cellular autoinducer molecules to communicate messages between cells and coordinate the response of the whole population (Henke and Bassler, 2004, Bassler and Losick, 2006, Camilli and Bassler, 2006, Joint *et al.*, 2007). Once the extracellular autoinducer threshold concentration has been reached, a signal transduction cascade induces gene expression alterations and amplifies signal molecule synthesis, concurrently altering community activities (Henke and Bassler, 2004, Waters and Bassler, 2005, Williams, 2007). Although quorum sensing systems function similarly across prokaryotes, the exact inducers and mechanisms vary between species, and may be specifically

adapted to suit a distinct niche (Waters and Bassler, 2005). Most quorum sensing signals function extracellularly and are small organic molecules or peptides less than 1 kDa in size (Williams, 2007). Different quorum sensing systems have been adopted by Gram-negative and Gram-positive species, and hybrid quorum sensing systems have also been observed (Henke and Bassler, 2004, Williams, 2007). Across all these systems, the signal molecule and its cognate receptor will be specialised to facilitate intraspecies communication; interspecies and cross-kingdom communication also occurs, using different quorum-sensing signal molecules (Williams, 2007, Pottathil *et al.*, 2008, Esmaeilishirazifard *et al.*, 2018).

The autoinducer signalling molecules of Gram negative bacteria are acyl homoserine lactones (AHLs) comprised of a homoserine lactone moiety with distinct acyl side chains; these are synthesised from S-adenosyl methionine and fatty acyl carrier proteins by LuxI-type AHL synthases (Sturme *et al.*, 2002, Camilli and Bassler, 2006). Many AHLs are able to move across membranes and are detected in the cytoplasm by LuxR-type proteins (Miller and Bassler, 2001). The ligand-bound LuxR-AHL complex then binds to DNA promoter elements in order to activate the which up- or down-regulate of quorum-sensing mediated genes (Bassler and Losick, 2006).

Gram positive bacteria utilise oligopeptide autoinducer peptides (AIPs) for quorum sensing (Sturme *et al.*, 2002, Camilli and Bassler, 2006) (Figure 1.1.). These AIPs are synthesised ribosomally and undergo post-translational modifications into mature forms that are typically 5–17 amino acids long (Kleerebezem *et al.*, 1997, Camilli and Bassler, 2006). These post-translational modifications may be in the form of lactone and thiolactone rings, lanthionines, or isoprenyl groups (Sturme *et al.*, 2002, Waters and Bassler, 2005). The functioning of these AIPS is broadly presented in Figure 1.2.: Peptide signals require exporters to move across the membrane and autoinducer processing and modification is often associated with signal release (Waters and Bassler, 2005). When the extracellular concentration of AIPs increases to a threshold (i.e. when cell density is sufficiently high) the signal molecule binds to membrane-bound two-component signalling proteins coupled to an intracellular response regulator (Camilli and Bassler, 2006). Signal transduction occurs via a phosphorylation cascade to modulate a DNA-binding transcriptional regulatory protein (viz., response regulator) which

initiates an autoinduction loop to activate the desired genes (Sturme *et al.*, 2002, Waters and Bassler, 2005, Camilli and Bassler, 2006).

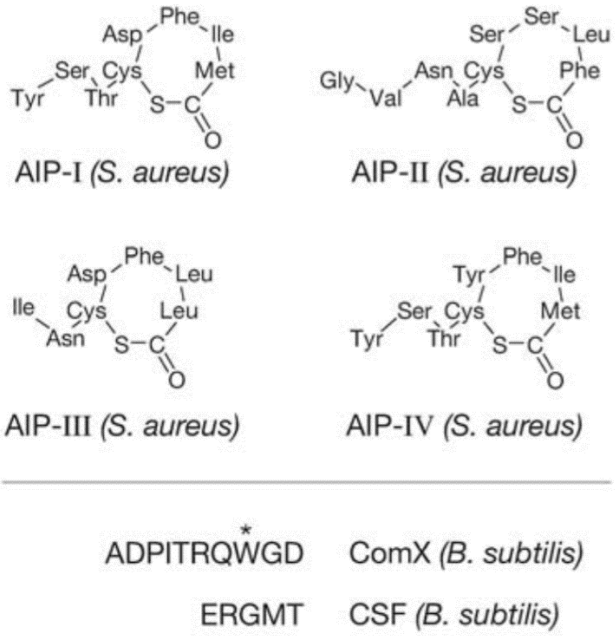


Figure 1.1. Representative autoinducer peptides used for Gram-positive bacterial quorum sensing (Waters and Bassler, 2005). *Staphylococcus aureus* utilises a thiolactone-ring-containing AIP, while *Bacillus subtilis* uses two forms of posttranslational modified AIP: The Competence and Sporulation Stimulating Factor (CSF) with a conserved ERGMT sequence, and the ComX pheromone which is variable between populations. ComX has a conserved and modified tryptophan residue being common to all variants and the asterisk over the *Bacillus subtills* ComX AIP denotes a post-translational modification on a tryptophan residue.

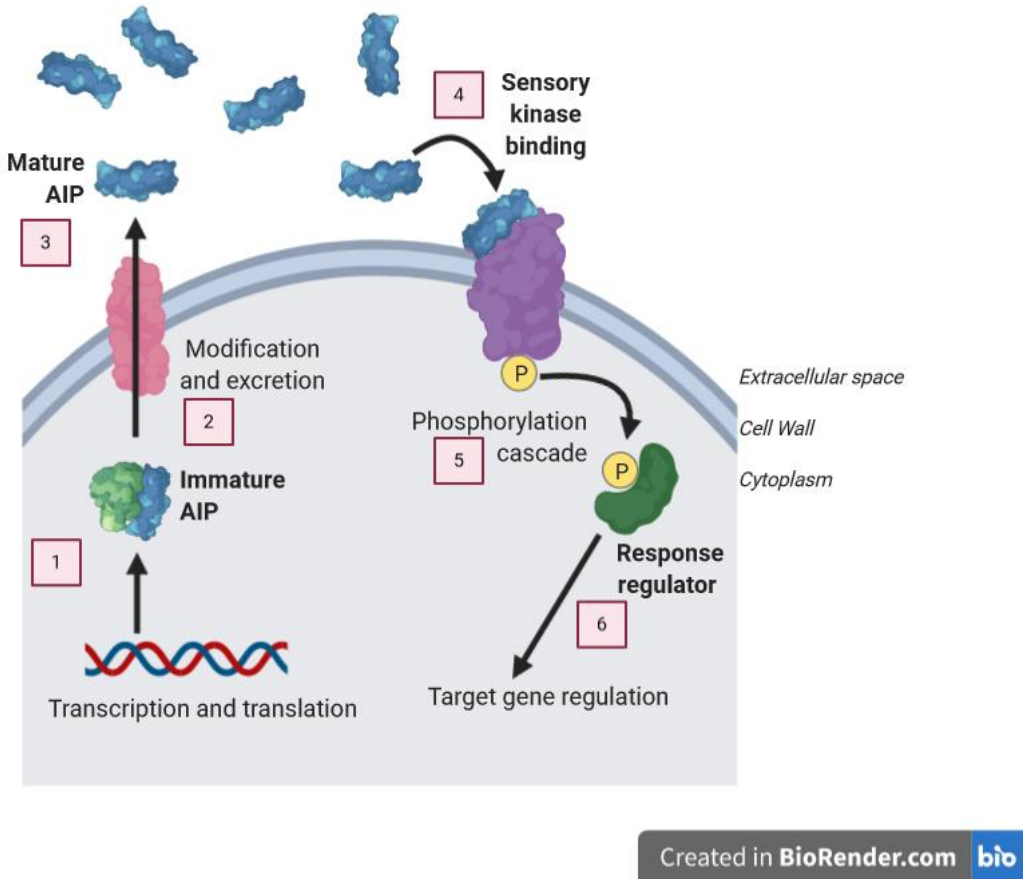


Figure 1.2. Gram-positive model of autoinducer peptide-mediated signalling cassette involved in quorum sensing. The immature AIP is transcribed and translated in the cell (1); the immature AIP undergoes simultaneous post-translational modification and secretion into the extracellular space (2); once the extracellular threshold concentrations of mature AIPs is reached (3), binding with the membrane-spanning sensor kinase is initiated (4); binding induces a phosphorylation cascade in the cell (5) which activates the response regulator and in turn activates quorum-sensing-mediated genes (6), and ultimately modulates population behaviours (Created with BioRender.com).

1.3. Plant-associated biofilms and plant-associated *Bacillus* spp.

Biofilm formation on a plant host has been described for many pathogenic and beneficial plant-associated bacterial species (Rudrappa *et al.*, 2008, Govindasamy *et al.*, 2010). Among other functions, quorum sensing by plant-associated bacteria allows for their spread of into distinct

niches by initiating microcolony formation, which will eventually develop into mature biofilms, and potentially into multispecies aggregations (Rudrappa *et al.*, 2008, Angus and Hirsch, 2013). Deposition onto plant surfaces may occur passively (e.g. wind and rain splash in the phyllosphere or water flow in the rhizosphere) or actively (e.g. motility and chemotaxis) (Ramey *et al.*, 2004). Plant-microbe interactions require direct physical contact between the host and the bacteria, and biofilms may form on the leaves, roots, seeds, and within the plant internal vasculature (Ramey *et al.*, 2004). Bacterial colonisation of the plant is heterogeneous owing to variations and fluctuations in water availability, nutrient substrates, and pH gradients, which impact bacterial swarming to favourable areas (Ramey *et al.*, 2004, Rudrappa *et al.*, 2008). Plant-derived compounds also play significant roles in the biofilm formation process, with 10–40% of a plant's photosynthates being released as root exudates (Rudrappa *et al.*, 2008). The *Bacillus*-host relationship is further reinforced by the secretion of certain plant exudates which act to specifically stimulate desirable strains to form biofilms (Yaryura *et al.*, 2008, Nihorimbere *et al.*, 2012, Beauregard *et al.*, 2013).

Plant-associated biofilms impact the plant in a variety of ways, notable examples include: virulence plasmid horizontal transfer in *Agrobacterium* spp.; regulation of nodule formation and transfer of plasmids encoding legume-specific nodulation in *Rhizobium* spp.; and, regulation of cell-wall degrading enzyme synthesis and production of secondary metabolites in pathogenic *Erwinia* spp. (Joint *et al.*, 2007). Biofilm formation in beneficial plant-associated *Bacillus* spp. has attracted significant research attention for the role that biofilms play in facilitating bacterial establishment and colonisation of the plant host (Bais *et al.*, 2004, Fan *et al.*, 2012, Chen *et al.*, 2013, Krober *et al.*, 2016, Yu *et al.*, 2016). Biofilms underpin the persistence of *Bacillus* spp. in the plant environment and are necessary for beneficial strains to carry out plant-health-promoting and disease antagonism activities, such as: plant growth hormone synthesis, extracellular enzyme production, and antimicrobial compound synthesis (Nagórksa *et al.*, 2007, Dogsa *et al.*, 2014). Species of *Bacillus* which have gained prominence as plant-health-promoting bacteria include members of the *Bacillus subtilis sensu lato* group—in particular *B. subtilis*, *B. amyloliquefaciens,* and *B. velezensis*—and strains of these species have evolved to meet the unique challenges of the plant-associated habitat (Zhang *et al.*, 2016). *Bacillus velezensis* strain FZB42 (formerly *B. amyloliquefaciens* subsp. *plantarum* FZB42) (Chen *et al.*, 2007) has gained significant attention over the last decade as it demonstrates

remarkable capabilities in plant-health maintenance and plant host colonisation (Chen *et al.*, 2009a, Chen *et al.*, 2009b, Fan *et al.*, 2011, Reva *et al.*, 2019). Strains of plant-associated *B. velezensis* were reported to have almost 400 genes geared towards maintaining the relationships with the plant host (Reva *et al.*, 2019) and, as a result, *B. velezensis* strain FZB42 has recently become the model strain for Gram-positive plant-health-promoting rhizobacteria (Fan *et al.*, 2019).

1.4. *Bacillus* spp. biofilms

Bacillus subtilis is the model species for Gram positive biofilm formation as it is capable of forming both submerged biofilms in liquid culture and biofilms at the agar-air and air-liquid interfaces (Vlamakis *et al.*, 2013, Dogsa *et al.*, 2014). *Bacillus* spp. are known to use quorum sensing to control many functions, including synthesis of antimicrobial peptides, sporulation, competence, and fruiting body formation (González-Pastor, 2007, Stefanic and Mandić-Mulec, 2009). Quorum sensing also serves to regulate the two stationary-phase cell types in *Bacillus;* namely naturally competent cells that are able to take in DNA from the environment, and endospore formers that are able to survive harsh environmental conditions in a dormant state (Stefanic and Mandić-Mulec, 2009).

The biofilm formation process in *B. subtilis* is initiated with the expression of matrix-production genes in response to extracellular signals, such as surfactin concentration, nutrient depletion, low oxygen, and surface adherence (Vlamakis *et al.*, 2013, Mielich-Suss and Lopez, 2015) (Figure 1.3.). Cells alter their lifestyle from a planktonic motile state to a sessile state by the simultaneous down-regulation of flagellar motility genes and up-regulation of ECM production genes (Mielich-Suss and Lopez, 2015). In the final stage of the biofilm life cycle, aerial projections arise from the biofilm surface; these fruiting bodies function as sites of sporulation and spore release (Branda *et al.*, 2001). Under laboratory conditions, the biofilm has a limited lifespan and will eventually disassemble and release endospores into the environment (Vlamakis *et al.*, 2013).

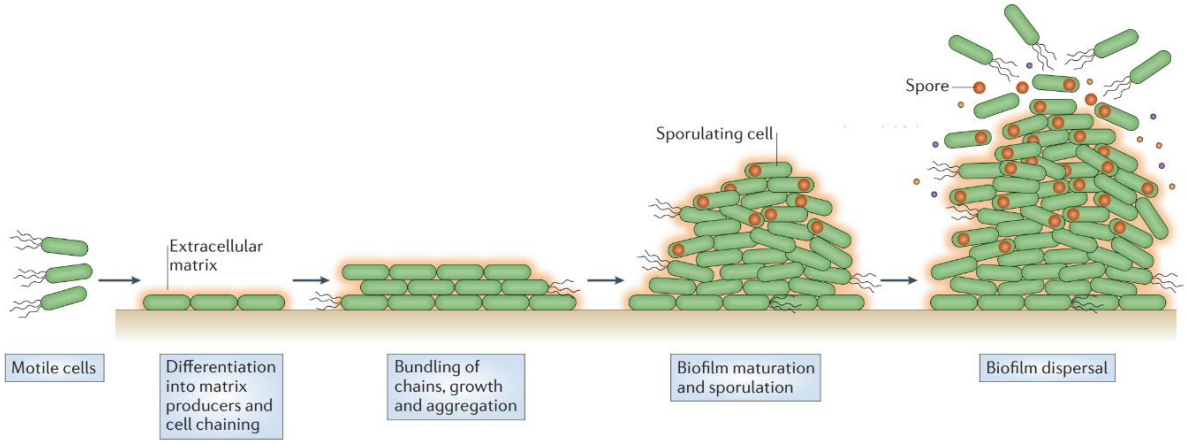


Figure 1.3. The formation, maturation, and disassembly of a *Bacillus subtilis* biofilm (adapted from Vlamakis *et al.*, 2013). Motile cells become sessile begin to aggregate, and secrete an ECM. With biofilm maturation over time, matrix-producing cells differentiate into sporulating cells and form aerial projections from the biofilm surface, from which spores are ultimately released into the environment, while other cells become motile and disperse from the biofilm.

The ECM provides the biofilm with a hydrophobic surface which is resistant to aqueous liquids and organic solvents (Vlamakis *et al.*, 2013). The foundation of the *Bacillus* spp. ECM is comprised of an EPS layer and proteins TasA and TapA (Branda *et al.*, 2006, Vlamakis *et al.*, 2013). TasA is a functional amyloid protein secreted into the extracellular space aided by SipW, where it assembles into fibres which are anchored to the cell wall by TapA (Romero *et al.*, 2010, Romero *et al.*, 2011). A third recently-described contributor, BslA, ensures biofilm integrity in the later phases of maturation (Kobayashi and Iwano, 2012). The hydrophobin protein BslA forms a protective hydrophobic layer over the biofilm and is important for complex colony morphology and pellicle formation (Kobayashi and Iwano, 2012, Hobley *et al.*, 2013). As the biofilm matures, cell clusters enlarge and become functionally differentiated, and so the community becomes organised within the protected matrix (Vlamakis *et al.*, 2013). The differentiated cell subpopulations are distinguishable at the phenotypic level but are genetically identical; in addition, each subtype responds to different signals and carries out specific functions within the community as a form of labour division and reduction of energy costs (Kearns, 2008).

1.4.1. Biofilm social differentiation in *Bacillus* spp.

Biofilm subpopulations come about from a combination of quorum sensing signals and stochastic events to stimulate functionally distinct cell types which coexist to perform different functions in the community (Kearns, 2008, Vlamakis *et al.*, 2013). This phenomenon is known as bimodality, whereby only a fraction of cells will differentiate into a certain subtype, each exhibiting distinct gene expression patterns (Dubnau and Losick, 2006). The regulatory network governing biofilm cell differentiation is regulated by specific signals, and the network is modulated to promote certain subpopulations depending on prevailing environmental conditions (Lopez and Kolter, 2010).

Bacillus subtilis cell differentiation into functional subtypes is controlled by three master regulators: Spo0A (Gaur *et al.*, 1986), DegU (Murray *et al.*, 2009), and ComA (Roggiani and Dubnau, 1993). The phosphorylated forms of the master regulators function to activate specific gene expression cascades to induce cell differentiation, and all three are required for stable robust biofilm formation and maintenance (Lopez and Kolter, 2010). When the regulators are in an unphosphorylated state the population remains planktonic and motile, until extracellular cues bring about motility loss and subpopulation differentiation (Mielich-Suss and Lopez, 2015). The cell subtypes within a *Bacillus* spp. biofilm are summarised as follows (Lopez and Kolter, 2010, Mielich-Suss and Lopez, 2015):-

- Motile cells: Enable biofilm expansion.
- Matrix producers: Synthesise ECM components.
- Cannibals: Provide nutrients and delay sporulation by initiating cell death.
- Spore-formers (derived from cannibal cells): Form endospores when conditions are unfavourable.
- Naturally competent cells: Take up exogenous DNA.
- Miners: Degrade large extracellular polymers into smaller nutritive units.

Subtype organisation in the *Bacillus* biofilm occurs in a predictable manner, though intersection in master-regulator-induced behaviours mean that some cells may overlap in their differentiated subtypes (Vlamakis *et al.*, 2008, Mielich-Suss and Lopez, 2015). The competent cells and surfactin-producers exist as small subpopulations throughout the biofilm; persistent

motile cells locate at the base of the biofilm to contribute to biofilm expansion; matrix-producers differentiate at the biofilm core to produce matrix and maintain biofilm rigidity; endospore formers localise at the aerial portions of the biofilm; and miners differentiate from motile or matrix-producer cells and exist in close proximity to the air interface for exoenzyme release and BlsA production (Mielich-Suss and Lopez, 2015).

The master regulators determining and controlling each subtype may result in an overlap in subtype functionality (Mielich-Suss and Lopez, 2015). Subtype differentiation in *Bacillus* occurs in three broad sequential steps (Lopez and Kolter, 2010). During exponential growth, a large fraction of cells is motile and the quorum-sensing signals (viz., ComX and surfactin) are at levels insufficient to trigger population responses. In stationary phase, the ComX concentration becomes sufficient to trigger differentiation of surfactin producers and competent cells. Increased surfactin production is sensed by neighbouring cells who differentiate into cannibals and matrix producers and, at this stage, the proportion of motile cells drops. Sporulating cells and miners are later differentiated in response to nutrient depletion and starvation-related events (Lopez and Kolter, 2010).

The master regulators, their mediating kinases and phosphatases, and the genes that they regulate are part of a vast network of inter-connected pathways operating in *Bacillus* spp. and some aspects of these regulator-mediated pathways may overlap (Schultz *et al.*, 2009) and link to other functions of *Bacillus* spp. such as competence development (Hamoen *et al.*, 2003b). Once activated by phosphorylation, each regulator directs discrete functions (summarised in Figure 1.4.): Spo0A~P controls a cascade for matrix production and regulates cannibalism and sporulation in a concentration-dependent manner; DegU~P controls protease synthesis and secretion; and ComA~P controls surfactin production (Lopez and Kolter, 2010). Metabolites secreted by other organisms may also impact the master regulators and influence differentiation, which allows *Bacillus* to adapt to the environment and to a specific environmental niche and to compete amongst other extant microbes (Lopez and Kolter, 2010).

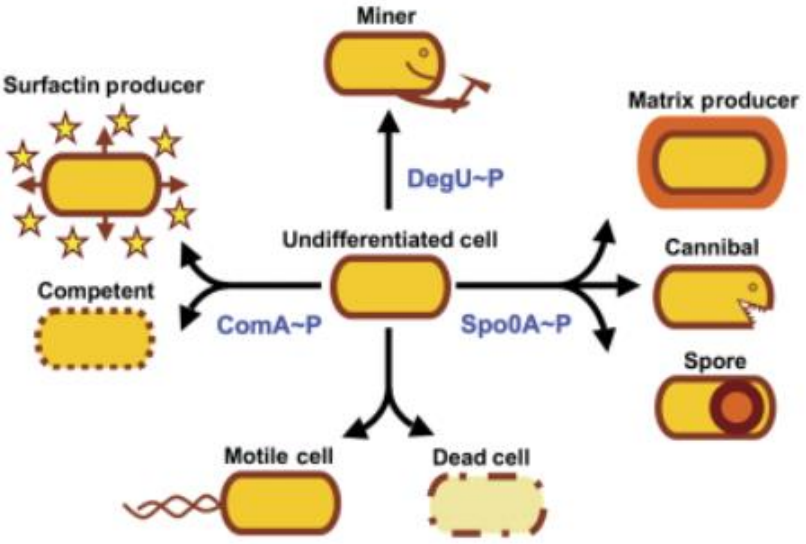


Figure 1.4. Summary of subtype differentiation observed in the *Bacillus subtilis* biofilm (Lopez and Kolter, 2010). Subtype differentiation and functioning within the biofilm is regulated by the phosphorylation of three master regulators: DegU, ComA, and Spo0A. Regulators are coupled to specific controlling phosphatases and kinases, which interaction to up- and down-regulate various genes involved in biofilm formation, maturation, and dispersal. These pathways are interconnected to others governing functions such as nutrient scavenging and competence development.

Spo0A is a central transcriptional regulator controlling more than one hundred genes, including those required for biofilm matrix gene expression and sporulation (Vlamakis *et al.*, 2013, Mielich-Suss and Lopez, 2015). Five sensor kinases (KinA–E) control Spo0A activation by the transfer of phosphoryl groups to Spo0A via a phosphorelay system, with each kinase catalysing various reactions and behaviours (Vlamakis *et al.*, 2013, Mielich-Suss and Lopez, 2015). The functions of spore formation, cannibalism, and production of ECM are overlapped and controlled by the phosphorylated form of Spo0A (Spo0A~P) (Vlamakis *et al.*, 2013). The concentration of Spo0A~P determines gene expression profile of that cell, and there are essentially two distinct levels of activation: lower levels which induce matrix production and cannibalism, and higher levels which induce sporulation (Fujita *et al.*, 2005). Cells sensitive to Spo0A~P are also cannibal cells, which secrete two toxins during low levels of Spo0A~P: Sporulation delaying protein (SdpC) and Sporulation killing factor (SkfA) (González-Pastor,

2011). These toxins do not affect cannibal cells but kill non-resistant neighbouring cells (including non-*B. subtilis* cells in mixed culture) (González-Pastor, 2007). The cannibalised cells are broken down into nutritive subunits by the miner subtype cells and thereby generate a food source to delay sporulation during extracellular nutrient scarcity (Mielich-Suss and Lopez, 2015).

The second master regulator DegU and its cytoplasmic kinase DegS, form a system which controls biofilm formation, genetic competence, motility, degradative enzyme production, and poly-gamma-glutamic acid production (Murray *et al.*, 2009). DegU acts as both a gene expression inhibitor and activator depending on its phosphorylation status (Mielich-Suss and Lopez, 2015). The activated form (DegU~P) creates a subpopulation of exoenzyme secretors (miners) which stimulate the degradation of large biopolymers into smaller nutritive forms for the community (Mielich-Suss and Lopez, 2015). Miners secrete enzymes such as the proteases subtilisin and bacillopeptidase, and polysaccharide-degrading levansucrase (Msadek, 1999). DegU~P regulation is controlled by DegS and associated phosphatases responding to amino acid starvation (Msadek *et al.*, 1990). Unphosphorylated DegU also plays a role in competence development, as it binds to the *com*K promoter to prevent competence; DegU~P is necessary for degradative enzyme production and positively regulates the production of amphiphilic BsIA protein (Kobayashi and Iwano, 2012, Hobley *et al.*, 2013).

Activation of the third master regulator ComA induces cells to synthesise surfactin, thereby triggering matrix production and the differentiation of surfactin producers and competent cells (Magnuson *et al.*, 1994, Mielich-Suss and Lopez, 2015). ComA is activated when extracellular ComX binds to its cognate membrane-associated histidine kinase ComP, whereafter ComP phosphorylates ComA (Magnuson *et al.*, 1994) (Figure 1.5.). ComA~P stimulates the expression of genes for surfactin biosynthesis within the *srfABCD* operon, and also the *com*S gene, which controls competence development, causing an overlap in cell function when transitioning between these two types (Marahiel *et al.*, 1993). ComS inhibits the proteolytic degradation of ComK, a transcriptional activator controlling the expression of structural genes for DNA uptake (Solomon and Grossman, 1996, Miller and Bassler, 2001).

The actions of master regulators are controlled by dephosphorylation to counteract kinase activity (Lopez and Kolter, 2010). The master regulator cascades are all negatively controlled intracellularly by proteins in the Rap (Response-regulator aspartyl-phosphate phosphatases) family (Pottathil and Lazazzera, 2003). The Rap family comprises eleven inhibitory proteins (RapA through to RapK) which are aspartyl-phosphate phosphatases that act by specifically dephosphorylating—and hence inhibiting—each of the master regulators (Mielich-Suss and Lopez, 2015). Rap-driven dephosphorylation directly, or indirectly, accelerates dephosphorylation of an active (~P) regulator, or may block the activation of a master regulator by binding to it and consequently blocking DNA binding by the regulator (Lopez and Kolter, 2010). Rap phosphatate activity is in turn controlled by a Rap-Phr cassette mechanism where individual Rap activity is mediated by a cognate extracellular Phr peptide, which binds to its intracellular Rap phosphatate when internalised (Pottathil and Lazazzera, 2003). Many of these Rap-Phr-mediated mechanisms are poorly understood owing to the complex natures of the different Rap-dependent genetic cascades (Auchtung *et al.*, 2006).

1.4.2. Autoinducers for *Bacillus subtilis* sporulation and competence development

Members of the *Bacillus subtilis* group of related taxa utilise the convergent pathways of two AIPs in their quorum-sensing-controlled response to high population density: a peptide pheromone, ComX, and a Phr peptide known as PhrC or Competence and Sporulation Stimulating Factor (CSF) (Magnuson *et al.*, 1994, Solomon *et al.*, 1995, Solomon *et al.*, 1996). Together, these two AIPs control the cellular levels of master regulator ComA~P and, as a result, the transcription of genes governing competence and sporulation (Solomon *et al.*, 1996, Dunny and Leonard, 1997, Tortosa and Dubnau, 1999, Henke and Bassler, 2004) (Figure 1.5.). In addition to their roles in competence development, these AIPs are believed to contribute to differing levels of *Bacillus* spp. communication: The CSF pentapeptide (ERGMT) is highly conserved and is considered to be involved in species-wide communication (Pottathil *et al.*, 2008); whereas the highly variable ComX pheromone is attributed to discrete population- and strain-level communication (Tortosa *et al.*, 2001).

As mentioned previously (Section 1.3.1.), the ComA response regulator induces surfactin synthesis and the overlapping cell subtypes of surfactin producers and competent cells (Auchtung *et al.*, 2006). As the population increases and environmental stresses become greater, some cells will commit to sporulation (Schultz *et al.*, 2009). During this process, the competence master regulator ComK is activated, and sporulating cells may become competent and take advantage of the DNA matter released by the cell lysis (Hamoen *et al.*, 2003a). The development of genetic competence is one example of a quorum-sensing-controlled function in *Bacillus* spp. which enables the cell to take up DNA from the environment via natural transformation (Sturme *et al.*, 2002, Hamoen *et al.*, 2003b). Competence is a reversible state initiated during a small window in the transition between logarithmic and stationary phases of growth, but will only occur in 10–20% of cells (Hamoen *et al.*, 2003b). Exogenous DNA is taken up by competent cells and assimilated into the bacterial genome to increase genetic variability within the community and for endogenous DNA repair (Tortosa and Dubnau, 1999). *Bacillus subtilis* competence machinery exhibits non-specific DNA uptake, and competent cells will take up plasmid, phage, and chromosomal DNA (Hamoen *et al.*, 2003b).

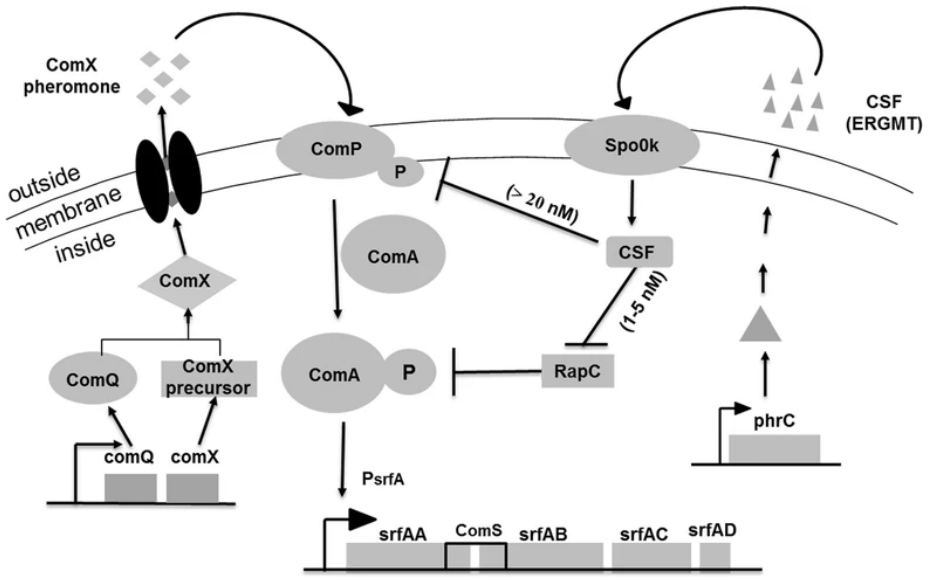


Figure 1.5. The convergent regulatory pathways of ComX and CSF in *Bacillus subtilis* (Guan *et al.*, 2015). Both pathways converge at response regulator ComA: ComX machinery functions to activate ComA by phosphorylation, and thereby upregulate competence and other functions. In contrast, CSF ultimately serves to antagonise active ComA and downregulate these functions.

1.4.2.1. ComX synthesis and function

The *B. subtilis* ComP-ComA two-component regulatory system is governed by the *comQXPA* gene loci, which encode the ComX peptide pheromone and the associated machinery (Tran *et al.*, 2000, Tortosa *et al.*, 2001). The *Bacillus comQXPA* gene loci encoding the peptide signalling cassette comprises *comQ* gene encoding isoprenyl transferase ComQ; *comX* encoding peptide pheromone ComX; *comP* encoding sensor kinase ComP; and *comA* encoding response regulator ComA (Tortosa *et al.*, 2001, Oslizlo *et al.*, 2014). Peptide signalling cassettes are found in many Gram positive bacteria, and genetically, a cassette encodes the precursor of the extracellular peptide immediately upstream of a kinase-response regulator gene pair, and the gene immediately upstream of the precursor encodes the machinery for any modification that may be required (Lazazzera *et al.*, 1999). In the *comQXPA* gene loci *comP* and *comA* genes are both found downstream of *comX*, and *comQ* immediately upstream of *comX* (Magnuson *et al.*, 1994, Lazazzera *et al.*, 1999).

The immature form of the ComX pheromone is transcribed as a peptide comprising 55 amino acids, which subsequently undergoes post-translational modification by ComQ to become the mature form that is 5–10 amino acids in length (Bacon Schneider *et al.*, 2002, Dogsa *et al.*, 2014). The post-translational modification of ComX involves an isoprenoid modification, which involves the addition of a hydrophobic group to a conserved tryptophan residue located three amino acids from the pre-ComX C-terminus (Dunny and Leonard, 1997, Tortosa *et al.*, 2001, Mandić-Mulec *et al.*, 2003). The processing adds ~200 Da to the peptide mass and increases its hydrophobicity, and is the only known incidence of an isoprenoidal tryptophan modification (Ansaldi *et al.*, 2002, Okada, 2011). Mature ComX is released from the cell and accumulates extracellularly until a threshold concentration is reached, where it then binds to its cognate ComP receptor on the cell surface (Piazza *et al.*, 1999, Dogsa *et al.*, 2014).

ComP is a histidine kinase signal transduction protein comprised of an amino-terminal membrane-associated domain of 6–8 membrane-spanning units, and a carboxy-terminal cytoplasmic domain containing conserved regions (Weinrauch *et al.*, 1990, Tortosa and Dubnau, 1999). The membrane-associated domain contains two extracellular loops, comprised

of approximately eighty aminoacyl residues (Weinrauch *et al.*, 1990). Once bound to ComX, the ComP undergoes autophosphorylation on a conserved histidine residue, and the phosphoryl group is transferred to an aspartate located on ComA (Lazazzera *et al.*, 1999, Dogsa *et al.*, 2014) (Figure 1.6.). As mentioned previously (Section 1.3.1), ComA~P stimulates the expression of several genes, including those for surfactin biosynthesis and competence development (Stefanic and Mandić-Mulec, 2009, Oslizlo *et al.*, 2014).

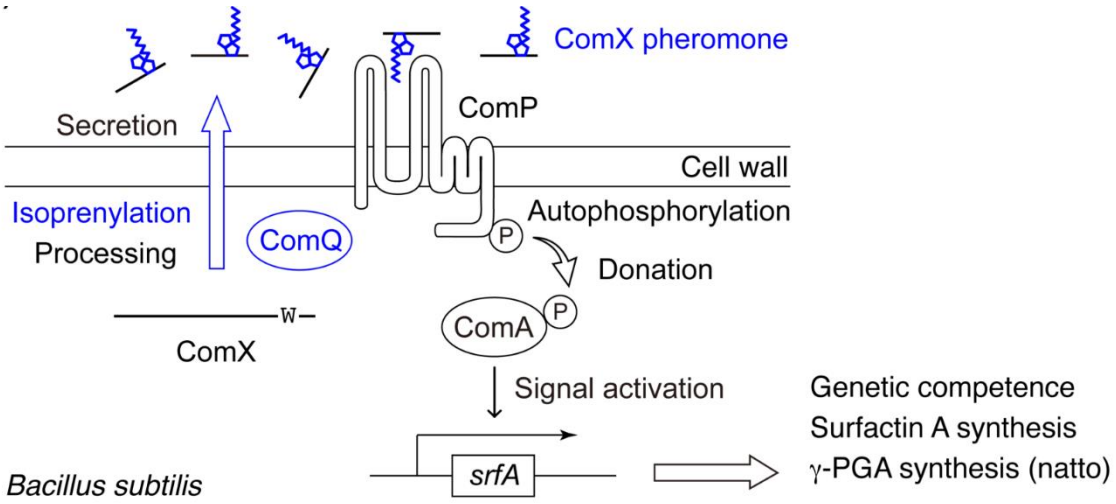


Figure 1.6. The pathways of ComX signalling in *Bacillus subtilis* (Okada *et al.*, 2017). Immature ComX undergoes simultaneous post-translational modification and secretion by ComQ; extracellular mature ComX binds to the ComP receptor on the cell surface to initiate a phosphorylation cascade, with the phosphorylation of response regulator ComA and the activation of the *srfA* gene which plays roles in such processes as genetic competence and surfactin synthesis.

1.4.2.2. ComX variants and pherotypes

Pheromone variations restrict communication within related groupings and are found in several quorum-sensing systems of Gram positive bacterial species (e.g. *Staphylococcus aureus*, *Streptococcus pneumoniae*, *B. subtilis*, and *B. cereus*) (Dogsa *et al.*, 2014). Social groupings arising from diversification in specialised communication pheromones are termed pherotypes (Stefanic *et al.*, 2012). The amino acid composition and nature of the post-translational

isoprenylation in the ComX peptide is highly variable between populations and has formed the basis for the distinction of pherotype groupings in *Bacillus* spp. (Okada *et al.*, 2005, Pottathil *et al.*, 2008). Distinct pherotypes within the *B. subtilis* group of related taxa have been identified from the ComX variants of strains *B. licheniformis* NCIMB-8874, *B. subtilis* 168, *B. subtilis* natto NAF4, *B. subtilis* RO-E-2, *B. subtilis* RS-D-2, *B. subtilis* RO-C-2 , *B. subtilis* RO-FF-1 , *B. subtilis* RS-B-1, *B. mojavensis* RO-B-2, *B. mojavensis* RO-H-1, and *B. mojavensis* RO-C-2 (Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002, Okada *et al.*, 2015, Esmaeilishirazifard *et al.*, 2017).

Common to all variants of ComX associated with *Bacillus* spp. is the tryptophan residue which undergoes an isoprenoid modification to become the mature peptide form (Okada *et al.*, 2008) (Figure 1.7.). Mass spectrometric analysis of ComX variants has determined that isoprenoid modifications increases the ComX mass through the addition of a farnesyl or geranyl group, resulting in the formation of a tricyclic ring structure (Okada, 2011). These tryptophan residue modifications mean that ComX pheromones have highly similar core structures, but differ in the lengths of the side chain (Okada, 2011). Hence, the length of the isoprenoid side chain is considered to be a greater determinant of pherotype-specificity than the ComX peptide sequence (Okada, 2011). Furthermore, ComX group-specificity has been linked to the nature of the tryptophan modification (Okada, 2011). For example: farnesyl-modified ComX variants elicit responses in each other, but not to the geranyl-modified pherotypes, and *vice versa*. Examples of amino acid sequences and posttranslational tryptophan modifications of several recognised ComX pherotypes are illustrated in Figure 1.7.: $ComX_{168}$ and $ComX_{RO\text{-}C\text{-}2}$ both carry farnesyl modifications, and are capable of inducing a greater degree of response in each other than to the other geranyl-modified groups $ComX_{RO\text{-}E\text{-}2,}$ $ComX_{RO\text{-}H\text{-}1,}$ $ComX_{RS\text{-}B\text{-}1,}$ and $ComX_{RO\text{-}B\text{-}2}$ (Okada, 2011). Mandić-Mulec *et al*. (2003) report ComX cross-reactivity amongst pherotypes. In their study, the authors report cross reactivity amongst the following pherotypes: 168 and RO-C-2, RO-A-4 and RTO-E-2, RO-FF-1 and DV3-E-3, and RS-D-2 and NAF4; while pherotypes RO-B-2 and RO-H-1 did not match to any of the other pherotypes studied.

The variable nature of *B. subtilis* ComX is evidenced at the genetic level. Analyses of the *comQXP* gene sequences from *Bacillus* spp. found the genes of *comQ*, *comX*, and N-terminal

comP sequences were sufficiently related to suggest co-evolution of these portions of the cassette (Ansaldi *et al.*, 2002, Ansaldi and Dubnau, 2004). Polymorphisms in the *comQXP* gene loci extending from *com*Q to *com*X and the N-terminal region of *comP* possess ~56% nucleotide similarity (Tran *et al.*, 2000, Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002, Ansaldi and Dubnau, 2004). Yet the portions of the cassette not directly involved with ComX reception—the C-terminus of *comP* and *comA*—are more conserved and exhibit >90% nucleotide similarity (Tran *et al.*, 2000, Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002). Furthermore, *comQXP* sequences are more variable than those of the surrounding genes and have a lower average G+C content than the *B. subtilis* genome (Ansaldi and Dubnau, 2004). This may be due to this region being under a stronger selection pressure, being subject to rapid mutation mechanisms, or having been obtained via horizontal gene transfer (Tortosa and Dubnau, 1999, Ansaldi and Dubnau, 2004).

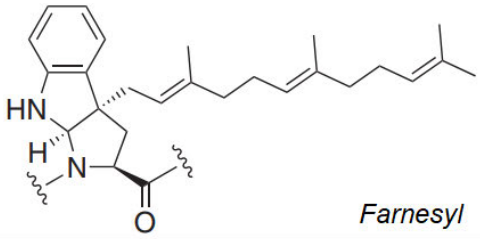
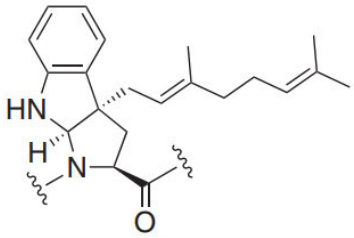
Bacillus strain	Amino acid sequence	Chemical structure of **W**
168	ADPITRQ**W**GD	 *Farnesyl*
RO-C-2	TRE**W**DG	
RO-E-2	GIF**W**EQ	 *Geranyl*
RO-H-1	MLD**W**KY	
RS-B-1	MMD**W**HY	
RO-B-2	YTNGN**W**VPS	

Figure 1.7. *Bacillus subtilis* ComX pheromone variants displaying the two types of post-translational modifications (Okada, 2011). The isoprenyl modification occurs on a conserved tryptophan residue (indicated in boldface "W"), and occur in one of two forms: $ComX_{168}$ and $ComX_{RO\text{-}C\text{-}2}$ have a farnesyl modification, while those variants below the line display geranyl modifications.

1.4.2.3. Competence and Sporulation Stimulating Factor (CSF) synthesis and function

CSF is the second AIP involved in *Bacillus* spp. competence development (Solomon and Grossman, 1996, Logan and De Vos, 2015). This AIP functions to antagonise the phosphorylated form of ComA in complement to ComX phosphorylation of this regulator (Solomon *et al.*, 1995, Henke and Bassler, 2004) (Figure 1.5.). CSF is controlled by several transcription factors which also have bearing on competence and sporulation (viz., Spo0A, AbrB, and Spo0H) (Solomon *et al.*, 1995). CSF (or PhrC) is a Phr phosphatase peptide encoded by the *phr*C gene, which is transcribed during the entry to the stationary growth phase when there are high levels of an RNA polymerase containing the alternate sigma factor (σH) (Lazazzera and Grossman, 1998, Lazazzera *et al.*, 1999). The transcribed forty-amino acid PhrC is processed into the mature CSF form by cleavage of five amino acids from the C-terminus (Lanigan-Gerdes *et al.*, 2007). The cleaved portion forms the mature CSF pentapeptide, glu-arg-gly-met-thr (ERGMT), which is then secreted extracellularly (Solomon *et al.*, 1995). Unlike ComX, CSF structure is conserved amongst *Bacillus* spp., and this may be due to its interaction with the equally conserved ComA regulator, and the role of CSF in *Bacillus* interspecies communication (Pottathil *et al.*, 2008).

Extracellular CSF functions intracellularly when it is transported back into the cell by an ABC-type oligopeptide transporter designated Opp (previously referred to as Spo0K) (Solomon *et al.*, 1996). The intracellular function of CSF is to inhibit cytoplasmic phosphatases, but its exact activities are dependent on its intracellular concentrations (Lazazzera and Grossman, 1998). CSF has demonstrated at least three roles in regulating ComA (Solomon *et al.*, 1995, Solomon *et al.*, 1996, Miller and Bassler, 2001):

- At low internal concentrations (0.2–5 nM) CSF stimulates expression of ComA-activated genes and inhibits cognate RapC;
- At higher internal concentrations (>20 nM) CSF inhibits expression of *srfA*; and,
- CSF stimulates sporulation by inhibiting RapB and inhibits *comS* expression by inhibiting ComP binding to ComA, serving to decrease competence in favour of sporulation.

1.5. Microbial speciation, pherotypes, and ecotypes

The prokaryotic domains of the tree of life display astonishing diversity at morphological and metabolic levels and, as such, microbial species were originally defined on the basis of these physiological traits (Ward, 2002, Ward *et al.*, 2007). Expansion in molecular techniques has seen microbial taxonomy advance from the purely culture-based phenotypic—and highly limited—view of microbial diversity afforded in the 19th and 20th centuries (Ward, 2002). Identifying microbial species is now benchmarked by the gene sequences of their small subunit rRNAs (viz., 16S rRNA subunit gene) (Cole *et al.*, 2010). Expanding bioinformatics capabilities and increased accessibility to gene sequencing technologies continue to increase the volume and range of molecular data available, and have deepened the resolution of genomic differences constituting microbial subspecies, strain, and population level diversity (Ward, 2006).

1.5.1. The microbial species concept

Species are the fundamental units and cataloguing cornerstones of microbial taxonomy, and understanding the microbial species concept is important in its application to many aspects of human activities (Cole *et al.*, 2010). A species definition exists for the purposes of medicine, biotechnology, and defence, where the goal is to formulate a set of simple and well-defined rules by which to define and name organisms that are similar in genomic and/or phenotypic characteristics (Doolittle and Papke, 2006). Current approaches to classifying Bacteria and Archaea apply a polyphasic operation-based model combining phenotypic characteristics, chemotaxonomic traits, and genotypic data and phylogenetics (Schleifer, 2009a).

Bacterial species are defined as being highly-similar genomically coherent groups delineated by standardised testing parameters (Gevers *et al.*, 2005). Operational-based classifications of bacteria concern phenotypic characteristics and genotypic aspects (Schleifer, 2009a). The genotypic "gold standard" thresholds of bacterial species demarcation apply several conventions: 60–70% similarity cut-off for DNA-DNA hybridisation (Stackebrandt, 2006); 98.65 % similarity in the 16S rRNA gene sequence (Kim *et al.*, 2014); and ~94% average

nucleotide identity (ANI) of genomes (Konstantinidis and Tiedje, 2005). If such criteria were to be applied to the primates (including humans), all known species would be classified into only one or two species (Ward, 2002); this fact highlights that prokaryotes are not amenable to species delineations as applied to macroorganisms (i.e. plants and animals) (Ward *et al.*, 2007).

Gene flow in prokaryotes (e.g. horizontal gene transfer) makes descent lineage more complicated to establish and so phenotypic characteristics have been retained for taxonomic descriptions (Schleifer, 2009a). Yet, in spite of advances in the resolution of microbial taxonomy, adherence to phenotypic-based naming conventions has resulted in bacterial strains with high-diversity genome content being categorised within one species (Ward, 2002). When applied to community-level studies, the phenotypic approach can underestimate functional diversity, structure, and community composition (Ward *et al.*, 2007). The disparate conventions of phenotypic and phylogenetic taxonomy have ultimately eroded the global framework for microbial species description and brought about the rise of polyphasic approaches to taxonomy by combining phylogenetic data to extant phenotypic criteria (Ward, 2006).

Phenotypic-based clustering has created a situation whereby a species may have been categorised on the basis of commonality within a set of core genes for vital abilities, yet members can harbour divergence in other genes which are susceptible to high variability (e.g. horizontal gene transfer) (Ward *et al.*, 2007). Prokaryote diversity is driven by many mechanisms, which typically include mobile elements (viz., prophages, plasmids, and transposons), immigration, speciation, horizontal (also known as lateral) gene transfer, and extinction events (Mandić-Mulec *et al.*, 2003, Cole *et al.*, 2010). While the essential coding regions conventionally applied in bacterial identification are subject to high levels of conservation (e.g. 16S rRNA gene), there remains ample opportunity for genetic exchange events in non-essential portions (Cole *et al.*, 2010).

Such variation in genome content and its associated determination of functional diversity resonates with the concepts of species genome (Lan and Reeves, 2000) or pangenome (Medini

et al., 2005). The pangenome concept proposes that a species be defined on genome-wide variation on the basis of core housekeeping genes (Medini *et al.*, 2005). The concept also acknowledges an accessory genome comprising the remaining diverse set of genes that a species may harbour which underpin such aspects as niche differentiation and functional divergence (Lan and Reeves, 2000). Variations in character or lifestyle genes which contribute to colonisation, survival, and adaptation to particular environment, niche, or habitat (Schleifer, 2009aa). The ubiquitous *Bacillaceae* family of bacteria are a prime example of variation in genes which impact diverse behaviour and environmental functioning; for example, among the tens of thousands of gene families described across 19 *Bacillus* species, only 814 orthologous genes were common to all of the genomes analysed (Alcaraz *et al.*, 2010). The variation in gene content amongst *Bacillus* spp. as a result of ecological specialisation has brought forward the relevance of ecology in systematics (Koeppel *et al.*, 2008).

The taxonomic debate continues as to the exact criteria applicable when defining a bacterial species (Ward, 2006, Ward *et al.*, 2007). The microbial species concept has become more than the simple need to name microorganisms, as it encompasses the composition, structure, and function of individual species within a community (Koeppel *et al.*, 2008). Different microbial populations and habitats are subject to different selection pressures and opportunities for genetic exchange. Yet, the drivers of speciation in microorganisms have not been fully elucidated, to the point where it may be argued that these drivers are variable between–and unique to–individual prokaryotic populations (Ward *et al.*, 2007). The paucity of a standardised microbial species concept which fully accounts for molecular diversity and functional phenotypic diversity means that microbial species and taxon definition is subjective, depending on which aspect of molecular variation the individual researcher prioritises, and the way in which these inter- and intra-variations are incorporated into naming convention (e.g. pangenome, core genome, species genome, functional taxa, and operational groups). The vigorous ongoing debate surrounding the microbial species concept and qualifying of the fundamental units of diversity within prokaryote communities has contributed several new concepts to microbial ecology, including an appreciation of ecologically distinct populations within a species (Cohan, 2005).

1.5.2. The bacterial ecotype

There are effectively two views of species demarcation in prokaryotic species: (1) The current convention of species as defined by traditional phenotypic traits and phylogenetically related groups, or that (2) species are dynamic and influenced by ever-changing evolutionary forces and genetic variations. The concept of ecological diversity in prokaryotes has been proposed to account for the intraspecies variation in bacterial genomes and to accommodate these within existing taxonomic convention (Cohan, 2006, Koeppel *et al.*, 2008). Ecological gradients in the molecular diversity of populations have been observed in several prokaryotic species, with population variants being distributed over different geographical locations, and combined with the additional observation of patterns of evolutionary relatedness amongst the variants in a single habitat (Ward, 2006).

The investigation of specialisation of microbial populations along ecological (Ecotype) or geographical (Geotype) scales seeks to differentiate the basic building blocks of a community structure which occupy different niches and fluctuate based on environmental variability (Ward, 2006). Population genetics approaches have determined that presently-described bacterial species may be ecologically distinct at the subpopulation level and may therefore be considered as "ecologically adapted species populations" or ecotypes (Cohan, 2005, Koeppel *et al.*, 2013). Ecotypes may be considered as fundamental units of diversity; and the ecological species concept could be key to appreciating a given microbial community, in terms of composition, structure, and function as an entire community (Ward, 2006, Koeppel *et al.*, 2008). Cohan (2005) proposes that ecological diversity in bacteria can arise via three mechanisms:

1) Niche-invasion mutations: a new ecotype arises from a new genotype which survives the selective purging of the ancestral ecotype.
2) Periodic selection mutations: events which purge diversity in an ecotype serve to increase divergence between ecotypes (i.e. purging-events within, but not between, ecotypes).
3) Speciation-quashing mutations: one ecotype extinguishes the other; an ecotype is rendered extinct even after selection events within the ecotype (e.g. an ecotype sharing

resources with another may be extinguished by a fitter adaptive mutant of the other ecotype).

An ecotype is formally defined as "a group of bacteria that are ecologically similar to one another, so similar that genetic diversity within the ecotype is limited by a cohesive force, either periodic selection or genetic drift, or both" (Cohan and Perry, 2007). Bacteria can alter their ecology by genetic changes: (1) mutations in extant genes, (2) allele transfer via homologous recombination, and (3) acquisition of new genes via horizontal gene transfer/plasmid/phage genetic material (Cohan, 2017). Genes which afford a complex adaptation can be acquired within one event, resulting in two subpopulations of a common gene pool but with different lifestyles (Doolittle and Papke, 2006). Cohan (2005) outlines an ecotype as a population of cells in the same ecological niche which can be out-competed to extinction by any adaptive mutant arising within the population, but the adaptive mutant cannot bring about the extinction of strains from other ecotypes. This contribution of various means of bacterial genetic exchange to the evolutionary process may then constitute a basis for microbial speciation (Koeppel *et al.*, 2008).

Eco(geo)types are frequently more closely related than the widely-applied divergence parameters based on 16S rRNA or DDH may suggest, with some variants maintaining entire similarity in the 16S rRNA gene sequence (Ward, 2006). For this reason, high-resolution DNA sequence analyses are necessary to better detect subtle divergences in sequence relationships. Gene sequence clusters, as determined by a variety of criteria, are expected to correspond to ecotypes, and sequence-based approaches suggest that a typically named species may comprise many ecotypes (Cohan, 2005). Interpreting the wealth of phylogenetic information in an ecological sense is made more ambiguous when considering factors such as geography, genetic drift, plasmid material, and rapid speciation, which often fail to correlate sequence divergence with ecological distinctness (Koeppel *et al.*, 2008). Ecotypes are commonly phylogenetically differentiated on the basis of sequence clusters of housekeeping genes (Koeppel *et al.*, 2008, Dogsa *et al.*, 2014), with the most common approaches being multilocus sequence approaches (Maiden, 2006) and Ecotype Simulation analyses (Cohan, 2002, Koeppel *et al.*, 2008) (both concepts are explored further in Section 1.4.2.).

The question remains as to just how ecotype diversification occurs within a species. Several models of speciation have been put forward to aid in the understanding and explanation of the evolution of the prokaryote species and the diversification into ecotypes (Presented in further detail in Cohan and Perry, 2007 and Ward *et al.*, 2007). These various models individually consider the means by which microbial diversity arises (i.e. mutation, recombination, or gene transfer) and the influence of the environment on these factors (e.g. genetic drift, purifying selection, positive selection, geographic isolation, and recombination) (Cohan and Perry, 2007, Ward *et al.*, 2007).

The Stable Ecotype Model (Illustrated in Figure 1.8.) is frequently favoured to describe microbial ecotypes, and presents a model in which the cohesive force of the species population is periodic selection (Cohan and Perry, 2007, Koeppel *et al.*, 2008). Population diversity originates from mutation or recombination within a population; although recombination between populations is thought to be too rarely-occurring to be a limiting factor in adaptive divergence between ecologically distinct populations (Ward *et al.*, 2007). The biodiversity in an ecosystem is regulated by periodic selection events—such as a change in environmental parameters—selects for a superior variant and reduces diversity within the population (Cohan, 2005; Cohan, 2006). As a result, one mutant/recombinant has greater fitness in that niche and outcompetes the original parental genome from the population, and the surviving variant then passes fitness traits onto the progeny, and further diversifies until the next periodic selection event (Cohan, 2017). As Cohan (2005) explains, the levels of sequence diversity observed in a bacterial species are usually too low to be the result of genetic drift and are therefore ascribed to periodic selection. Recombination events in prokaryotes are rare, and selection events will favour an adaptive mutant, thereby diminishing the diversity within an ecotype, and ensuring that diversity will only persist until the next selection event. Periodic selection then seeks to explain ecotype variation by its purging of diversity within, but not between, ecotypes.

Bacterial strains which cluster over the majority of genome content, but differ markedly in smaller gene regions, are considered to have undergone a form of periodic selection known as "adapt globally, act locally" mutations (Cohan, 2005). In this particular periodic selection model, the adaptive mutation can recombine with other ecotypes; when the adaptive mutation

is transferred into a new ecotype a selection event is initiated within this ecotype (Cohan, 2005). Such lineage divergence involves the transfer of part of the original lineage into a new niche, which is a microhabitat of new resources and/or novel prevailing conditions, and may extend even to a microhabitat within the same locale (Cohan, 2017). On occasion, an ecological variant may advance into a new niche and is then unaffected by periodic selection for the niche occupied by its parents (Ward, 2006). In this case, a new population is formed, which is subject to its private selection events; in turn causing the two populations to further diverge with each event (Cohan, 2017) (Figure 1.8.).

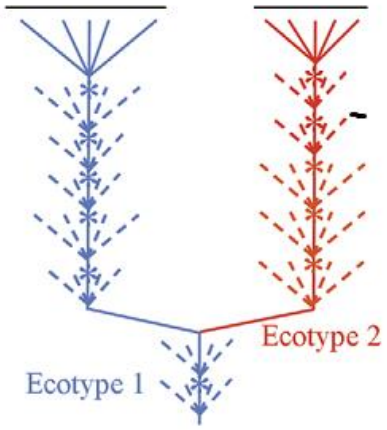


Figure 1.8. The Stable Ecotype Model as defined by Cohan and Perry (2007) and Ward *et al.* (2007). Periodic selection limits the diversity within an ecologically distinct population in a particular niche. Niche invasion mutations in a population (Blue) create a new ecologically distinct variant which founds a new population (Red). Each population undergoes discrete mutations and is subject to periodic selection during its lifetime (dotted lines), eventually resulting in separate phylogenetic clusters which represent distinct ecotypes (viz., Ecotype 1 and Ecotype 2).

The Stable Ecotype Model assumes periodic selection to occur multiple times in the lifetime of the ecotype, and that eventually periodic selections and random genetic events within each separate population will result in genetically distinct phylogenetic clusters, interpreted as ecologically unique species or ecotypes (Cohan, 2006). Ecological divergence can aid the emergence of new species when the lineage carries species attributes of being ecologically

distinct, when there exists cohesion "within but not between different species", and when species are distinguishable as gene sequence clusters (Cohan, 2017). Bacterial cohesion arises from periodic selection where the two lineages diverge permanently, and ultimately manifest as species clusters. Irreversible separateness arises when these species can coexist indefinitely, which is necessary for new genera/family/species/phyla emergence (Cohan, 2017). Precisely how to incorporate the ecotype concept into naming convention remains unresolved. For practical purposes, Koeppel *et al.* (2008) put forward that ecotypes be denoted by an ‘‘ecovar’’ epithet added to the existing species name.

In summary, ecotypes arise from adaptive mutants which diversify to colonise novel microhabitats and niches. New ecotypes are formed infrequently, but an ecotype will undergo many periodic selection events during its existence (Cohan, 2005). Sequence diversification between ecotypes does occur and sequence clustering in these systems is dominated by periodic selection, therefore allowing parallels between ecotypes and sequence clusters in genomic analyses (Koeppel *et al.*, 2008).

1.5.3. Techniques for defining ecotypes in bacteria

Not all bacterial species and strains descend from clonal lineages, because mechanisms of genetic exchange play a significant role in prokaryote species divergence (Ibarz Pavon and Maiden, 2009). Factors such as large population sizes, frequency of genetic exchange, small size of recombined segments, and the relative simplicity of their physiology allow bacteria to evolve rapidly in terms of new species emergence (cladogenesis) and niche adaptability (anagenesis) (Koeppel *et al.*, 2013). Conventional bacterial taxonomy relies on the standard of 16S rRNA gene sequence relationships, but this gene is considered to be insufficiently sensitive to resolve subspecies-, strain-, and population-level divergences and may be distorted by homologous recombination (Hanage *et al.*, 2006, Ward *et al.*, 2007). Several high-resolution analyses have been applied to define ecotypes: ANI (Meier-Kolthoff *et al.*, 2013), average amino acid identity (AAI) (Richter and Rossello-Mora, 2009), digital DNA-DNA hybridisation (dDDH) (Konstantinidis and Tiedje, 2005), and tetranucleotide signatures (TETRA) (Teeling *et al.*, 2004), and multilocus sequence typing (MLST) (Maiden, 2006). Ecotypes may be

differentiated on the basis of phylogenetic sub-clusters in housekeeping genes (Cohan, 2006, Koeppel *et al.*, 2008, Dogsa *et al.*, 2014), and so MLSA will be examined as an ecotyping approach.

MLST was originally developed for analysing *Neisseria meningitides* isolates, as an alternative to applying the conventionally-used small subunit ribosomal RNA gene sequence analysis and large-dataset whole-genome sequencing approaches (Maiden *et al.*, 1998). MLST was accepted as a standardisable and universally applicable technique for the molecular characterisation of bacterial isolates and has been expanded as MLSA for other prokaryotic species and for broader use in molecular epidemiology, population biology, and evolutionary biology studies (Maiden, 2006, Ibarz Pavon and Maiden, 2009). Initial MLST studies evaluated restriction fragments of target gene loci (hence MLSTyping), though modern variations rely on phylogenetic analysis (hence MLSAnalysis) of multiple protein-encoding sequence targets (Cole *et al.*, 2010). Because MLSA evaluates differences in the same gene targets between different species or populations, it is also used in new species description and to define relationships and evolutionary divergence between closely-related genera and species (Hanage *et al.*, 2006). The advances in gene and genome sequencing, and the expansion of internet-based genetic databases have seen MLSA become widely used as a reproducible, fast, and universally-applicable means of characterising bacterial strains and isolates (Ibarz Pavon and Maiden, 2009).

In essence, the MLSA method targets genetic polymorphisms within a subset of carefully-selected housekeeping gene loci using phylogenetic analysis (Maiden, 2006). The MLSA method is highly flexible between studies, and there are several aspects which should be considered when designing an MLSA which, according to Glaeser and Kämpfer (2015), should include numbers of gene targets, sequence quality, gene fragment lengths, alignment choice, concatenation, and phylogeny model choice. Housekeeping genes are applied since the proteins they encode are necessary to life and will be found in all strains of a genus, and the use of multiple gene sequences in MLSA serves to offer a greater number of informative nucleotide sites and avoid biases from recombination at single loci (Hanage *et al.*, 2006). Loci selection is flexible between individual MLSA studies but remains an important aspect, specifically in

the case of intrapopulation diversity studies. Loci should be selected on the basis of several criteria; essentially they must be universal amongst the taxa of interest, contain highly conserved regions for primer design, and not be linked within the genome (in order to limit the influences of horizontal gene transfer and recombination events) (Cole *et al.*, 2010, Glaeser and Kämpfer, 2015).

Over time, an ecotype will become genetically identifiable as a distinct sequence cluster, and appear as a monophyletic group in a DNA-based phylogeny (Cohan, 2005). However, caution must be exercised when using DNA sequence-based analyses to define ecotypes, since geographically-isolated populations which belong to the same ecotype can diverge into separate sequence clusters. This results when the adaptive mutant from one region has not had the opportunity to compete with populations from other regions, hence sequence divergence can occur between geographically-isolated populations of an ecotype (Cohan, 2005). Several algorithms have been derived in order to avoid subjectivity in sequence analysis for ecotype delineation. Commonly-used algorithms are now available as online tools for analysis, and include: Ecotype Simulation (ES) (Cohan and Perry, 2007, Koeppel *et al.*, 2008); AdaptML (Hunt *et al.*, 2008); Generalised Mixed Yule-Coalescent (GMYC) (Barraclough *et al.*, 2009); and, Bayesian Analysis of Population Structure (BAPS) (Corander and Tang, 2007). Francisco *et al.* (2014) evaluated each of these models for the purposes of ecotype derivation amongst *Bacillus* spp. and determined that ES was the best suited for analysing the models of habitat preference and the criteria of the Stable Ecotype model.

1.6. Taxonomy and ecotypes within the *Bacillus subtilis sensu lato* group

The genus *Bacillus* is currently largest genus in the *Bacillaceae*, with 226 species described (as of September 2014) (Mandić-Mulec *et al.*, 2015). Members of this genus have a broad range of metabolic and phenotypic characteristics, and possess substantial genetic diversity (Schleifer, 2009b). *Bacillus* phylogeny has traditionally been defined by the 16S rRNA gene sequence, which is able only to resolve low levels of divergence between species, and cannot accurately differentiate the range of phenotypic characteristics these species possess (Mandić-Mulec *et al.*, 2015). Distinguishing between *Bacillus* species on the basis of classical taxonomy

parameters (i.e. Morphological and physiological characteristics, cell wall composition, 16S rRNA sequence, G+C percentage, DNA-DNA hybridisation, and FAME) often leads to misallocations (Fan *et al.*, 2017). The multiple mechanisms of gene heterogeneity, and low discrimination afforded by the 16S rRNA gene sequence, has led to supplementary approaches being recommended for diversity assessment such as DNA-DNA hybridisation and the sequencing of additional core genes (e.g. *rpoB*, *cheA*, and *gyrA*) for *Bacillus* spp. characterisation (Reva *et al.*, 2004, Mandić-Mulec *et al.*, 2015).

The sharing of genetic material between members of the *Bacillaceae* occurs via mechanisms such as transposons, horizontal gene transfer, phage-mediated exchange, and conjugative plasmid transfer (Fajardo-Cavazos *et al.*, 2014). Consequently, *sensu lato* groupings have been named (e.g. *B. subtilis sensu lato* group) as a measure to account for these functionally-variable yet closely-related groups of taxa (Mandić-Mulec *et al.*, 2015). Two major *sensu lato* groupings have been described in the genus *Bacillus*: *B. cereus* and *B. subtilis* (Fritze, 2004). Members of these groups of related taxa are phylogenetically and phenotypically homogeneous and are frequently indistinguishable by 16S rRNA gene sequence phylogeny alone (Fritze, 2004, Rooney *et al.*, 2009). The *B. cereus sensu lato* group (as at 2003) is comprised of *B. cereus, B. anthracis, B. weihenstephanensis, B. mycoides, B. pseudomycoides,* and *B. thuringiensis*. The *B. subtilis sensu lato* group originally comprised *B. subtilis*, *B. amyloliquefaciens*, *B. licheniformis, and B. pumilus,* but has been significantly expanded and revised over the last decade.

The *B. subtilis* group of related taxa comprises those species most commonly present in the plant-associated environment and appreciated to possess plant-health-promotion activities, such as *B. subtilis, B. velezensis,* and *B. amyloliquefaciens,* and as such, will be the focus of this review. With many novel approaches allowing greater resolution of genetic heterogeneity (e.g. ANI, AAI, TETRA, and dDDH) multiple species have since been added to the *B. subtilis sensu lato* group: *B. atrophaeus, B. mojavensis, B. vallismortis, B. sonorensis, B. velezensis, B. axarquiensis, B. tequilensis, B. aerius, B. aerophilus, B. stratosphericus, B. altitudinis, B. safensis, B. methylotrophicus, B. siamensis, B. xiamenensis, B. vanillea, B. paralicheniformis, B. glycinifermentas, B. oryzicola, B. gobiensis,* and *B. nakamurai* (Fan *et al.*, 2017). A study

of 14 commercialised plant-health-product *B. subtilis sensu lato* strains found that many had been taxonomically misclassified (Dunlap, 2019). This is not uncommon in this bacterial family, and many of these species have undergone revisions and have been reassigned to other species counterparts (Fan *et al.*, 2017). For example: *B. subtilis sensu stricto* undergone revision to comprise three subspecies: *subtilis*, *spizizeni,* and *inaquosorum* (Nakamura *et al.*, 1999, Rooney *et al.*, 2009).

A notable example of the ecotype debate and its influence on taxonomy has concerned plant-associated *B. amyloliquefaciens* strains. The study presented by Reva *et al.* (2004) found evidence for ecotypes within plant-associated *B. amyloliquefaciens* on the basis of *cheA* and *gyrA* sequences. Divergence within this species had been taken further, with Borriss *et al.* (2011) proposing a plant-associated *plantarum* subspecies of *B. amyloliquefaciens.* The proposed type strain for *B. amyloliquefaciens* subsp. *plantarum* FZB42 (FZB42^{T}=DSM 23117^{T} =BGSC 10A6^{T}) has been the focus of much disagreement and reconsideration of the validity of this subspecies designation (Fan *et al.*, 2017). This type strain was subsequently determined to show a relationship to *B. velezensis* (Ruiz-Garcia *et al.*, 2005), *B. methylotrophicus* (Madhaiyan *et al.*, 2010), *B. oryzicola*, and *B. vanillea* type strains (Dunlap *et al.*, 2015, Dunlap *et al.*, 2016). The *B. amyloliquefaciens* subsp. *plantarum* subspecies was then later classified as a heterotypic synonym of *B. methylotrophicus*, or *B. velezensis* (Dunlap *et al.*, 2015, Dunlap *et al.*, 2016). Analyses using AAI, TETRA, *rpo*B gene sequence data, and mean G+C % have suggested that the conspecific group comprising *B. amyloliquefaciens* subsp. *plantarum*, *B. methylotrophicus*, and *B. velezensis* may simply be defined under the singular species *B. velezensis* (Dunlap *et al.*, 2016). The most recent taxonomic analyses further propose that the *B. amyloliquefaciens* clade be considered as a taxonomic unit above of species level, designated as an "operational group *B. amyloliquefaciens*" currently comprised of soil-borne *B. amyloliquefaciens*, and plant-associated *B. siamensis* and *B. velezensis* (Fan *et al.*, 2017). The latter two species have been encompassed as *B. velezensis* due to genomic adaptations and propensity for a plant-associated lifestyle, and the well-studied plant-associated *B. velezensis* strain FZB42 has recently been adopted as the model organism for Gram positive plant growth-promoting *Bacillus* spp. (Fan *et al.*, 2019).

The increasing resolution and capability of molecular techniques, has allowed the observation of phylogenetic groupings suggestive of genomic variability linked to ecological variation in multiple studies involving *Bacillus* spp.:

1) Geographical divergence in *B. subtilis* and *B. mojavensis* strains based on housekeeping gene RFLP (Roberts and Cohan, 1995);
2) Reva *et al.* (2004) assessed housekeeping genes variation amongst 17 environmental *Bacillus* strains, and found degrees of variance suggestive of adaptation to the plant-associated environment;
3) *Bacillus simplex* isolates from the Evolution Canyons based on observed microclimate adaptations, and MLSA and RAPD-PCR approaches (Sikorski and Nevo, 2005);
4) Death Valley *B. subtilis* and *B. licheniformis* ecotypes have been detected by ES and AdaptML algorithms (Connor *et al.*, 2010);
5) Adaptive mutations in *Bacillus* spp. which allow niche advancement and colonisation have been described by Koeppel *et al.* (2013);
6) Thermal tolerance adaptations of *B. thuringiensis* isolates identified by MLST and pulsed-field gel electrophoresis (Swiecicka *et al.*, 2013);
7) Ecotypes amongst marine *B. pumilus* have been described using MLSA (Liu *et al.*, 2013); and,
8) Ecological speciation amongst Death Valley *Bacillus subtilis* isolates has been defined using MLSA (Kopac *et al.*, 2014).

The above-mentioned studies illustrate that a polyphasic approach combining ES with MLSA, together with genomic fingerprinting approaches allows detection of putative ecotypes amongst populations of environmental *Bacillus* species. Such high-resolution analyses of the genetic relationships between environmental members within the *B. subtilis sensu lato* grouping suggests that diversification into ecologically distinct groupings (i.e. ecotypes) can occur other *B. subtilis sensu lato* member species. Given the ubiquity of *Bacillus* spp. in aquatic-, soil-, and plant-associated environments it is reasonable to expect that these organisms have adapted mechanisms to thrive in specific ecological niches and diversified into ecotype populations. Being able to detect and describe ecological variation may afford benefits

to human activities, such as being able to match a biocontrol agent to a host plant and locality's soil type and climate. It then becomes of interest to explore the extent of these adaptations and determine whether humankind could harness them in activities such as agriculture, pollution control, and soil sustainability.

1.7. Research problem: Relating plant-associated *Bacillus* ecotypes and pherotypes

From the ecotype perspective, the *sensu lato* groupings of *Bacillus* are a fitting example of how small gene variations can confound conventional taxonomy, particularly within the contested *B. velezensis*/*B. amyloliquefaciens* subsp. *plantarum* subgrouping, where plant-associated ecological adaptation at the genetic level impacts taxonomy (Fan *et al.*, 2017). Recognising the divergence of niche-adapted *Bacillus* subtypes (Ecotypes)—such as *B. velezensis* and forms of *B. subtilis*—which have evolved to coexist within the plant and soil environments is of particular worth when one considers the beneficial nature of these strains in terms of plant health and, ultimately, agricultural productivity (Govindasamy *et al.*, 2010).

One functional adaptation with niche-advancement implications for *B. subtilis sensu lato* is the ComX pheromone quorum sensing system and resultant pherotype groupings. The *B. subtilis sensu lato* species *comQXPA* gene loci demonstrate high intraspecies diversity and distinct pherotype groupings have been identified in the gene sequences within in this group (Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002). As such, a population would gain advantage my proliferating in a chosen niche, and protecting public goods by developing discrete communication pheromone variants. Research evidence supports the relationship between ecotype and pherotype in soil-derived *B. subtilis sensu lato* (Stefanic *et al.*, 2012). With specific focus on *Bacillus* spp., Stefanic and Mandić-Mulec (2009) examined the correlation between pherotypes (defined by variances in the *comQXPA* gene loci) and ecotypes within 1 cm^3 soil samples defined by ES. The study determined the existence of three ecotypes in their samples, which each demonstrating a distinct dominant pherotype. On the basis of these findings, Stefanic *et al.* (2012) further proposed ecotype-driven pherotype evolution, which may have come about as a means of isolating intra-ecotype communication. Pherotypes are highly variable amongst

any given *Bacillus* community, to the point that each habitat may harbour multiple pherotype variants (Stefanic and Mandić-Mulec, 2009, Oslizlo *et al.*, 2015).

Stefanic *et al.* (2012) further reported that there existed other low-frequency pherotypes within a single ecotype, and found that these pherotypes were shared with other ecotypes in the vicinity. One explanation for pherotype variability in an ecotype is presented by the "pherotype cycling" hypothesis, where a minority pherotype within an ecotype becomes advantaged by being at low cell density and not inducing its own quorum sensing response. Pherotype cycling results in the dominant pherotype in a community constantly evolving through the creation of new low-frequency pherotypes via the introduction of new genetic material during competence development (viz., horizontal gene transfer) (Eldar, 2011). The fact the pherotypes are shared between ecotypes (Stefanic and Mandić-Mulec, 2009) suggests that the evolution of novel pherotypes may be infrequent, possibly due to the need for preservation within the complex ComX-associated machinery. Stefanic and Mandić-Mulec (2009) suggest that although the pherotype cycling model drives the selection for unique pherotypes, the presence of a low number of pherotype forms would suggest that truly novel pherotypes appearing in a species may occur as a result of gene transfer.

Modelling of social conflict scenarios by Eldar (2011) suggests bacterial quorum-sensing systems evolve as a response to cheater populations taking advantage of public goods, and that this sets up pherotype cycling. Being "rare phenotype" population affords fitness benefit by allowing this "rare" pherotype to support itself by poaching public goods until it reaches cell densities sufficient to induce its own quorum sensing response (Stefanic *et al.*, 2012). That a specific ComX pheromone induces a quorum sensing response restricted to pherotype group members suggests that ComX-specific communication could endow the population with a competitive advantage by allowing it to take advantage of public goods and increase their number without contributing to the whole community (Ansaldi *et al.*, 2002, Bacon Schneider *et al.*, 2002, Stefanic *et al.*, 2012).

Ecotypes have been associated with distinct *Bacillus* spp. pherotypes, and so pherotype diversity may be viewed as an adaptation to ecological diversity in *B. subtilis*. Tortosa and Dubnau (1999) suggest that a deficiency of communication between natural isolates of the same species may function to lower the frequency of genetic exchange and is a means of increasing fitness, and so the variability and specificity in this quorum sensing system may then represent a mechanism of sexual isolation. Koeppel *et al.* (2013) agreed with the notion that "speciation is easy" in their study of *Bacillus* spp. divergence, by demonstrating that adaptive mutations leading to new ecotypes appeared to be about as common as those improving fitness within an existing ecotype. The evolution of such quorum-sensing specificity suggests that a population becomes highly adapted to its habitat, which in turn implies that a pherotype may also represent a form of ecological specialisation, much like an ecotype (Stefanic and Mandić-Mulec, 2009).

In spite of the various bodies of works examining *B. subtilis sensu lato* ecotypes and ComX variation in environmental populations, there is a paucity of directed studies as to the prevalence of these two concepts in the strictly plant-associated context. Oslizlo *et al.* (2015) examined pherotype diversity and some degree of housekeeping gene variation amongst *Bacillus* spp. from the tomato rhizoplane, but did not address the existence of ecotype variation amongst the strains evaluated. There is the distinct possibility that novel *B. subtilis sensu lato* pherotypes may exist amongst the ecotype populations in the plant-associated environment. Studies of *Bacillus* spp. pherotypes have been limited to several model pherotypes in B. *subtilis, B. mojavensis,* and *B. licheniformis* strains (Ansaldi and Dubnau, 2004, Esmaeilishirazifard *et al.*, 2017). Comparatively little exploration has been conducted on other *B. subtilis sensu lato* members and, to date, no research has explored the ComX diversity of plant-associated *B. velezensis* strains and ecotypes amongst this species specifically. If ecological strain variants possessed multiple pherotypes within a small volume of riverbank soil (Stefanic and Mandić-Mulec, 2009) then it stands to reason that a similar scenario could be evident in the plant-associated habitat in which multiple populations of a *Bacillus* spp. exist in competition with each other.

The ecotype and pherotype groups extant amongst plant-associated *Bacillus* spp. is of interest from both microbial ecology and biocontrol perspectives. A biocontrol agent introduced into the plant environment must be able to compete with the existing microflora and establish in its desired niche to carry out beneficial and disease-antagonism activities (Reva *et al.*, 2004, Yaryura *et al.*, 2008). In order to achieve this goal, the population must form a stable biofilm and colonise its niche; both of these goals are indirectly dependant on the synthesis of the peptide pheromone ComX. Knowledge of an ecotype's dominant pherotype may aid to promote biocontrol agent establishment. Promotion of beneficial strain colonisation could be achieved by approaches such as the co-addition of purified ComX elicitors to boost colonisation, or by matching an agent's ecotype to those extant on the plant to increase the natural population of desirable strains. Based on the evidence present in this chapter, the present study hypothesises that: (1) Plant-associated *B. subtilis sensu lato* isolates with biocontrol capabilities constitute locally-adapted South African ecotypes, and that (2) These ecotypes employ unique ComX variants. The present study aims to investigate the diversity of ecotypes and pherotypes amongst a subset of plant-associated wild-type isolates from the *B. subtilis sensu lato* group (chiefly *B. velezensis* and *B. subtilis*) using housekeeping gene sequences; and to explore the diversity within quorum-sensing genes (viz., *comQ* and *comQXP*). This information will be applied to determine whether there is any basis for an observable relationship between the ecotypes and pherotypes within the isolates evaluated, and the degree of pherotype variation between the model pherotypes and those extant in the South African context.

1.8. References

Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L. & Olmedo, G. (2010) Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* **11**: 332, 1–17. 10.1186/1471-2164-11-332

Angus, A. A. & Hirsch, A. M. (2013) Biofilm formation in the rhizosphere: Multispecies interactions and implications for plant growth. *Molecular Microbial Ecology of the Rhizosphere* Vol. 1 (de Bruijn, F. J., ed.) pp. 701–712. John Wiley & Sons, Inc., Hoboken, NJ, USA. 10.1002/9781118297674.ch66

Ansaldi, M. & Dubnau, D. (2004) Diversifying selection at the *Bacillus* quorum-sensing locus and determinants of modification specificity during synthesis of the ComX pheromone. *Journal of Bacteriology* **186**: 1, 15–21. 10.1128/jb.186.1.15-21.2004

Ansaldi, M., Marolt, D., Stebe, T., Mandić-Mulec, I. & Dubnau, D. (2002) Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Molecular Microbiology* **44**: 6, 1561–1573. https://doi.org/10.1046/j.1365-2958.2002.02977.x

Auchtung, J. M., Lee, C. A. & Grossman, A. D. (2006) Modulation of the ComA-dependent quorum response in *Bacillus subtilis* by multiple Rap proteins and Phr peptides. *Journal of Bacteriology* **188**: 14, 5273–5285. 10.1128/jb.00300-06

Bacon Schneider, K., Palmer, T. M. & Grossman, A. D. (2002) Characterization of *comQ* and *comX*, two genes required for production of ComX pheromone in *Bacillus subtilis*. *Journal of Bacteriology* **184**: 2, 410–419. 10.1128/JB.184.2.410-419.2002

Bais, H. P., Fall, R. & Vivanco, J. M. (2004) Biocontrol of *Bacillus subtilis* against infection of *Arabidopsis* roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiology* **134**: 1, 307–319. 10.1104/pp.103.028712

Barraclough, T. G., Hughes, M., Ashford-Hodges, N. & Fujisawa, T. (2009) Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biology Letters* **5**: 3, 425–428. 10.1098/rsbl.2009.0091

Bassler, B. L. & Losick, R. (2006) Bacterially speaking. *Cell* **125**: 2, 237–246. 10.1016/j.cell.2006.04.001.

Beauregard, P. B., Chai, Y., Vlamakis, H., Losick, R. & Kolter, R. (2013) *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 17, E1621–E1630. 10.1073/pnas.1218984110

Borriss, R. (2011) Use of plant-associated *Bacillus* strains as biofertilizers and biocontrol agents in agriculture. *Bacteria in Agrobiology: Plant Growth Responses* (Maheshwari, D. K., ed.) pp. 41–76. Springer-Verlag, Berlin, Germany. 10.1007/978-3-642-20332-9_3

Branda, S. S., Vik, S., Friedman, L. & Kolter, R. (2005) Biofilms: The matrix revisited. *Trends in Microbiology* **13**: 1, 20–26. 10.1016/j.tim.2004.11.006

Branda, S. S., González-Pastor, J. E., Ben-Yehuda, S., Losick, R. & Kolter, R. (2001) Fruiting body formation by *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 20, 11621–11626. 10.1073/pnas.191384198

Branda, S. S., Chu, F., Kearns, D. B., Losick, R. & Kolter, R. (2006) A major protein component of the *Bacillus subtilis* biofilm matrix. *Molecular Microbiology* **59**: 4, 1229–1238. 10.1111/j.1365-2958.2005.05020.x

Camilli, A. & Bassler, B. L. (2006) Bacterial small-molecule signaling pathways. *Science* **311**: 5764, 1113–1116. 10.1126/science.1121357

Cawoy, H., Bettiol, W., Fickers, P. & Ongena, M. (2011) *Bacillus*-based biological control of plant diseases. *Pesticides in the Modern World: Pesticides Use and Management* (Stoytcheva, M., ed.) pp. 273–302. InTech. 10.5772/17184

Chen, X. H., Koumoutsi, A., Scholz, R. & Borriss, R. (2009a) More than anticipated: Production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. *Journal of Molecular Microbiology and Biotechnology* **16**: 1–2, 14–24. 10.1159/000142891

Chen, X. H., Koumoutsi, A., Scholz, R., Schneider, K., Vater, J., Süssmuth, R., Piel, J. & Borriss, R. (2009b) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *Journal of Biotechnology* **140**: 27–37. 10.1016/j.jbiotec.2008.10.011

Chen, X. H., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W. R., Reva, O., Junge, H., Voigt, B., Jungblut, P. R., Vater, J., Süssmuth, R., Liesegang, H., Strittmatter, A., Gottschalk, G. & Borriss, R. (2007) Comparative analysis of the complete genome sequence of the plant growth–promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nature Biotechnology* **25**: 9, 1007–1014. 10.1038/nbt1325

Chen, Y., Yan, F., Chai, Y., Liu, H., Kolter, R., Losick, R. & Guo, J. H. (2013) Biocontrol of tomato wilt disease by *Bacillus subtilis* isolates from natural environments depends on conserved genes mediating biofilm formation. *Environmental Microbiology* **15**: 3, 848–864. 10.1111/j.1462-2920.2012.02860.x

Cohan, F. M. (2002) What are bacterial species? *Annual Reviews in Microbiology* **56**: 457–487. 10.1146/annurev.micro.56.012302.160634

Cohan, F. M. (2005) Periodic selection and ecological diversity in bacteria. *Selective Sweep* (Nurminsky, D., ed.) pp. 78–93. Springer US, Boston, MA. 10.1007/0-387-27651-3_7

Cohan, F. M. (2006) Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **361**: 1475, 1985–1996. 10.1098/rstb.2006.1918

Cohan, F. M. (2017) Transmission in the origins of bacterial diversity, from ecotypes to phyla. *Microbiology Spectrum* **5**: 5, 1–26. doi:10.1128/microbiolspec.MTBP-0014-2016

Cohan, F. M. & Perry, E. B. (2007) A systematics for discovering the fundamental units of bacterial diversity. *Current Biology* **17**: 10, 373–386. 10.1016/j.cub.2007.03.032

Cole, J. R., Konstandinidis, K., Farris, R. J. & Tiedje, J. M. (2010) Microbial diversity and phylogeny: Extending from rRNAs to genomes. *Environmental Molecular Microbiology* (Liu, W.-T. & Jansson, J. K., eds.), pp. 1–19. Caister Academic Press, Norfolk, UK.

Compant, S., Duffy, B., Nowak, J., Clement, C. & Barka, E. A. (2005) Use of plant growth-promoting bacteria for biocontrol of plant diseases: Principles, mechanisms of action, and future prospects. *Applied and Environmental Microbiology* **71**: 9, 4951–4959. 10.1128/AEM.71.9.4951-4959.2005

Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koeppel, A. F., Burger, A., Cole, S. G., Perry, E. B., Krizanc, D., Field, N. C., Slaton, M. & Cohan, F. M. (2010) Ecology of speciation in the genus *Bacillus*. *Applied and Environmental Microbiology* **76**: 5, 1349–1358. 10.1128/aem.01988-09

Corander, J. & Tang, J. (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**: 1, 19–31. https://doi.org/10.1016/j.mbs.2006.09.015

Dogsa, I., Oslizlo, A., Stefanic, P. & Mandić-Mulec, I. (2014) Social interactions and biofilm formation in *Bacillus subtilis*. *Food Technology and Biotechnology* **52**: 2, 149–157. 10.1128/JB.01290-08

Doolittle, W. F. & Papke, R. T. (2006) Genomics and the bacterial species problem. *Genome Biology* **7**: 9, 116.111–116.117. 10.1186/gb-2006-7-9-116

Dubnau, D. & Losick, R. (2006) Bistability in bacteria. *Molecular Microbiolgy* **61**: 3, 564–572. 10.1111/j.1365-2958.2006.05249.x

Dunlap, C. A. (2019) Taxonomy of registered *Bacillus* spp. strains used as plant pathogen antagonists. *Biological Control* **134**: 82–86. https://doi.org/10.1016/j.biocontrol.2019.04.011

Dunlap, C. A., Kim, S.-J., Kwon, S.-W. & Rooney, A. P. (2015) Phylogenomic analysis shows that *Bacillus amyloliquefaciens* subsp. *plantarum* is a later heterotypic synonym of *Bacillus methylotrophicus*. *International Journal of Systematic and Evolutionary Microbiology* **65**: 7, 2104–2109. doi:10.1099/ijs.0.000226

Dunlap, C. A., Kim, S.-J., Kwon, S.-W. & Rooney, A. P. (2016) *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and *'Bacillus oryzicola*' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. *International Journal of Systematic and Evolutionary Microbiology* **66**: 3, 1212–1217. doi:10.1099/ijsem.0.000858

Dunny, G. M. & Leonard, B. A. (1997) Cell-cell communication in Gram-positive bacteria. *Annual Review of Microbiology* **51**: 527–564. 10.1146/annurev.micro.51.1.527

Eldar, A. (2011) Social conflict drives the evolutionary divergence of quorum sensing. *Proceedings of the National Academy of Sciences Of the United States of America* **108**: 33, 13635–13640. 10.1073/pnas.1102923108

Esmaeilishirazifard, E., De Vizio, D., Moschos, S. A. & Keshavarz, T. (2017) Genomic and molecular characterization of a novel quorum sensing molecule in *Bacillus licheniformis*. *AMB Express* **7**: 78, 1–13. 10.1186/s13568-017-0381-6

Esmaeilishirazifard, E., Dariush, A., Moschos, S. A. & Keshavarz, T. (2018) A novel antifungal property for the *Bacillus licheniformis* ComX pheromone and its possible role in inter-kingdom cross-talk. *Applied Microbiology and Biotechnology* **102**: 12, 5197–5208. 10.1007/s00253-018-9004-7

Fajardo-Cavazos, P., Maughan, H. & Nicholson, W. L. (2014) Evolution in the *Bacillaceae*. *Microbiology Spectrum* **2**: 5, 1–32. 10.1128/microbiolspec.TBS-0020-2014

Fan, B., Borriss, R., Bleiss, W. & Wu, X. (2012) Gram-positive rhizobacterium *Bacillus amyloliquefaciens* FZB42 colonizes three types of plants in different patterns. *The Journal of Microbiology* **50**: 1, 38–44. 10.1007/s12275-012-1439-4

Fan, B., Blom, J., Klenk, H.-P. & Borriss, R. (2017) *Bacillus amyloliquefaciens*, *Bacillus velezensis* and *Bacillus siamensis* form an 'operational group *B. amyloliquefaciens*' within the *B. subtilis* species complex. *Frontiers in Microbiology* **8**: 22, 1–15. 10.3389/fmicb.2017.00022

Fan, B., Chen, X. H., Budiharjo, A., Bleiss, W., Vater, J. & Borriss, R. (2011) Efficient colonization of plant roots by the plant growth promoting bacterium *Bacillus amyloliquefaciens* FZB42, engineered to express green fluorescent protein. *Journal of Biotechnology* **151**: 4, 303–311. http://dx.doi.org/10.1016/j.jbiotec.2010.12.022

Fan, B., Wang, C., Ding, X., Zhu, B., Song, X. & Borriss, R. (2019) AmyloWiki: An integrated database for *Bacillus velezensis* FZB42, the model strain for plant growth-promoting *Bacilli*. *Database: The Journal of Biological Databases and Curation* **2019**: baz071. 10.1093/database/baz071

FAO, I., UNICEF, WFP, and WHO (2017) The state of food security and nutrition in the world 2017: Building resilience for peace and food security. pp. 1–27. FAO, Rome. 978-92-5-109888-2

Francisco, J. C., Cohan, F. M. & Krizanc, D. (2014) Accuracy and efficiency of algorithms for the demarcation of bacterial ecotypes from DNA sequence data. *International Journal of Bioinformatics Research and Applications* **10**: 4–5, 409–425. 10.1504/ijbra.2014.062992

Fritze, D. (2004) Taxonomy of the genus *Bacillus* and related genera: The aerobic endospore-forming bacteria. *Phytopathology* **94**: 11, 1245–1248. https://doi.org/10.1094/PHYTO.2004.94.11.1245

Fujita, M., González-Pastor, J. E. & Losick, R. (2005) High- and low-threshold genes in the Spo0A regulon of *Bacillus subtilis*. *Journal of Bacteriology* **187**: 4, 1357–1368. 10.1128/JB.187.4.1357-1368.2005

Guan, C., Cui, W., Cheng, J., Zhou, L., Guo, J., Hu, X., Xiao, G. & Zhou, Z. (2015) Construction and development of an auto-regulatory gene expression system in Bacillus subtilis. *Microbial Cell Factories* **14**: 1, 150. 10.1186/s12934-015-0341-2

Gaur, N. K., Dubnau, E. & Smith, I. (1986) Characterization of a cloned *Bacillus subtilis* gene that inhibits sporulation in multiple copies. *Journal of Bacteriology* **168**: 2, 860–869. 10.1128/jb.168.2.860-869.1986

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F. L. & Swings, J. (2005) Re-evaluating prokaryotic species. *Nature Reviews Microbiology* **3**: 9, 733–739. 10.1038/nrmicro1236

Glaeser, S. P. & Kämpfer, P. (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology* **38**: 4, 237–245. http://dx.doi.org/10.1016/j.syapm.2015.03.007

González-Pastor, J. E. (2007) Multicellularity and social behaviour in *Bacillus subtilis*. *Bacillus: Cellular and Molecular Biology* (Graumann, P., ed.) pp. 418–449. Caister Academic Press, Norfolk, UK.

González-Pastor, J. E. (2011) Cannibalism: A social behavior in sporulating *Bacillus subtilis*. *FEMS Microbiology Reviews* **35**: 3, 415–424. doi:10.1111/j.1574-6976.2010.00253.x

Govindasamy, V., Senthilkumar, M., Magheshwaran, V., Kumar, U., Bose, P., Sharma, V. & Annapurna, K. (2010) *Bacillus* and *Paenibacillus* spp.: Potential PGPR for sustainable agriculture. *Plant Growth and Health Promoting Bacteria* (Maheshwari, D. K., ed.) pp. 333–364. Springer-Verlag, Berlin, Germany. https://doi.org/10.1007/978-3-642-13612-2

Hamoen, L. W., Venema, G. & Kuipers, O. P. (2003b) Controlling competence in *Bacillus subtilis*: Shared use of regulators. *Microbiology* **149**: 9–17. 10.1099/mic.0.26003-0

Hamoen, L. W., Kausche, D., Marahiel, M. A., Sinderen, D., Venema, G. & Serror, P. (2003a) The *Bacillus subtilis* transition state regulator AbrB binds to the 35 promoter region of *comK*. *FEMS Microbiology Letters* **218**: 2, 299–304. 10.1111/j.1574-6968.2003.tb11532.x

Hanage, W. P., Fraser, C. & Spratt, B. G. (2006) Sequences, sequence clusters and bacterial species. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **361**: 1475, 1917–1927. 10.1098/rstb.2006.1917

Henke, J. M. & Bassler, B. L. (2004) Bacterial social engagements. *Trends in Cell Biology* **14**: 11, 648–656. 10.1016/j.tcb.2004.09.012

Hobley, L., Ostrowski, A., Rao, F. V., Bromley, K. M., Porter, M., Prescott, A. R., MacPhee, C. E., van Aalten, D. M. & Stanley-Wall, N. R. (2013) BslA is a self-assembling bacterial hydrophobin that coats the *Bacillus subtilis* biofilm. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 33, 13600–13605. 10.1073/pnas.1306390110

Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J. & Polz, M. F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 5879, 1081–1085. 10.1126/science.1157890

Ibarz Pavon, A. B. & Maiden, M. C. (2009) Multilocus sequence typing. *Molecular Epidemiology of Microorganisms Methods in Molecular Biology (Methods and Protocols)* Vol. 551 (Caugant, D., ed.) pp. 129–140. Humana Press, Totowa, NJ, USA. 10.1007/978-1-60327-999-4_11

Joint, I., Allan Downie, J. & Williams, P. (2007) Bacterial conversations: Talking, listening and eavesdropping. An introduction. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **362**: 1483, 1115–1117. 10.1098/rstb.2007.2038

Kearns, D. B. (2008) Division of labour during *Bacillus subtilis* biofilm formation. *Molecular Microbiology* **67**: 2, 229–231. 10.1111/j.1365-2958.2007.06053.x

Kim, M., Oh, H. S., Park, S. C. & Chun, J. (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* **64**: 346–351. 10.1099/ijs.0.059774-0

Kleerebezem, M., Quadri, L. E. N., Kuipers, O. P. & De Vos, W. M. (1997) Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Molecular Microbiology* **24**: 5, 895–904. 10.1046/j.1365-2958.1997.4251782.x

Kobayashi, K. & Iwano, M. (2012) BslA(YuaB) forms a hydrophobic layer on the surface of *Bacillus subtilis* biofilms. *Molecular Microbiology* **85**: 1, 51–66. 10.1111/j.1365-2958.2012.08094.x

Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., Rooney, A. P., Brambilla, E., Connor, N., Ratcliff, R. M., Nevo, E. & Cohan, F. M. (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 7, 2504–2509. 10.1073/pnas.0712205105

Koeppel, A. F., Wertheim, J. O., Barone, L., Gentile, N., Krizanc, D. & Cohan, F. M. (2013) Speedy speciation in a bacterial microcosm: New species can arise as frequently as adaptations within a species. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology* **7**: 6, 1080–1091. 10.1038/ismej.2013.3

Konstantinidis, K. T. & Tiedje, J. M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 7, 2567–2572. 10.1073/pnas.0409727102

Kopac, S., Wang, Z., Wiedenbeck, J., Sherry, J., Wu, M. & Cohan, F. M. (2014) Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. *Applied and Environmental Microbiology* **80**: 16, 4842–4853. 10.1128/aem.00576-14

Krober, M., Verwaaijen, B., Wibberg, D., Winkler, A., Puhler, A. & Schluter, A. (2016) Comparative transcriptome analysis of the biocontrol strain *Bacillus amyloliquefaciens* FZB42 as response to biofilm formation analyzed by RNA sequencing. *Journal of Biotechnology* **231**: 212–223. 10.1016/j.jbiotec.2016.06.013

Lan, R. & Reeves, P. R. (2000) Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends in Microbiology* **8**: 9, 396–401. 10.1016/S0966-842X(00)01791-1

Lanigan-Gerdes, S., Dooley, A. N., Faull, K. F. & Lazazzera, B. A. (2007) Identification of subtilisin, Epr and Vpr as enzymes that produce CSF, an extracellular signalling peptide of *Bacillus subtilis*. *Molecular Microbiology* **65**: 5, 1321–1333. 10.1111/j.1365-2958.2007.05869.x

Lazazzera, B. A. & Grossman, A. D. (1998) The ins and outs of peptide signaling. *Trends in Microbiology* **6**: 7, 288–294. https://doi.org/10.1016/S0966-842X(98)01313-4

Lazazzera, B. A., Kurtser, I. G., McQuade, R. S. & Grossman, A. D. (1999) An autoregulatory circuit affecting peptide signaling in *Bacillus subtilis*. *Journal of Bacteriology* **181**: 17, 5193–6200. 10.1128/JB.181.17.5193-5200.1999

Lemon, K. P., Earl, A. M., Vlamakis, H. C., Aguilar, C. & Kolter, R. (2008) Biofilm development with an emphasis on *Bacillus subtilis*. *Current Topics in Microbiology and Immunology* **322**: 1–16. 10.1007/978-3-540-75418-3_1

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G. & Shao, Z. (2013) Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. *PLoS One* **8**: 11, e80097. 10.1371/journal.pone.0080097

Logan, N. A. & De Vos, P. (2015) *Bacillus*. *Bergey's Manual of Systematics of Archaea and Bacteria* (Whitman, W. B., ed.) pp. 1–163. John Wiley & Sons / Bergey's Manual Trust. 10.1002/9781118960608.gbm00530

Lopez, D. & Kolter, R. (2010) Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. *FEMS Microbiology Reviews* **34**: 2, 134–149. 10.1111/j.1574-6976.2009.00199.x

Madhaiyan, M., Poonguzhali, S., Kwon, S. W. & Sa, T. M. (2010) *Bacillus methylotrophicus* sp. nov., a methanol-utilizing, plant-growth-promoting bacterium isolated from rice rhizosphere soil. *International Journal of Systematic and Evolutionary Microbiology* **60**: 2490–2495. 10.1099/ijs.0.015487-0

Magnuson, R., Solomon, J. & Grossman, A. D. (1994) Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell* **77**: 2, 207–216. 10.1016/0092-8674(94)90313-1

Maiden, M. C. (2006) Multilocus sequence typing of bacteria. *Annual Review of Microbiology* **60**: 561–588. 10.1146/annurev.micro.59.030804.121325

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M. & Spratt, B. G. (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 6, 3140–3145. 10.1073/pnas.95.6.3140

Mandić-Mulec, I., Stefanic, P. & van Elsas, J. D. (2015) Ecology of *Bacillaceae*. *Microbiology Spectrum* **3**: 1, 1–23. 10.1128/microbiolspec.TBS-0017-2013

Mandić-Mulec, I., Kraigher, B., Cepon, U. & Mahne, I. (2003) Variability of the quorum sensing system in natural isolates of *Bacillus* sp. *Food Technology and Biotechnology* **47**: 1, 23–28.

Marahiel, M. A., Nakano, M. M. & Zuber, P. (1993) Regulation of peptide antibiotic production in *Bacillus*. *Molecular Microbiology* **7**: 5, 631–636. 10.1111/j.1365-2958.1993.tb01154.x

Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. (2005) The microbial pan-genome. *Current Opinion in Genetics & Development* **15**: 6, 589–594. https://doi.org/10.1016/j.gde.2005.09.006

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**: 60. 10.1186/1471-2105-14-60

Mielich-Suss, B. & Lopez, D. (2015) Molecular mechanisms involved in *Bacillus subtilis* biofilm formation. *Environmental Microbiology* **17**: 3, 555–565. 10.1111/1462-2920.12527

Miller, M. B. & Bassler, B. L. (2001) Quorum sensing in bacteria. *Annual Review of Microbiology* **55**: 165–199. 10.1146/annurev.micro.55.1.165

Msadek, T. (1999) When the going gets tough: Survival strategies and environmental signaling networks in Bacillus subtilis. *Trends in Microbiology* **7**: 5, 201–207. https://doi.org/10.1016/S0966-842X(99)01479-1

Msadek, T., Kunst, F., Henner, D., Klier, A., Rapoport, G. & Dedonder, R. (1990) Signal transduction pathway controlling synthesis of a class of degradative enzymes in *Bacillus subtilis:* Expression of the regulatory genes and analysis of mutations in *deg*S and *deg*U. *Journal of Bacteriology* **172**: 2, 824–834. 10.1128/jb.172.2.824-834.1990

Murray, E. J., Kiley, T. B. & Stanley-Wall, N. R. (2009) A pivotal role for the response regulator DegU in controlling multicellular behaviour. *Microbiology* **155**: 1, 1–8. doi:10.1099/mic.0.023903-0

Murthy, P. S. & Venkatesan, R. (2009) Industrial biofilms and their control. *Marine and Industrial Biofouling* (Flemming, H.-C., Murthy, P. S., Venkatesan, R. & Cooksey, K., eds.), pp. 65–101. Springer, Berlin, Heidelberg. 10.1007/978-3-540-69796-1_4

Nagórksa, K., Bikowski, M. & Obuchowski, M. (2007) Multicellular behaviour and production of a wide variety of toxic substances support usage of *Bacillus subtilis* as a powerful biocontrol agent. *Acta Biochimica Polonica* **54**: 3, 495–508. 10.18388/abp.2007_3224

Nakamura, L. K., Roberts, M. S. & Cohan, F. M. (1999) Relationship of *Bacillus subtilis* clades associated with strains 168 and W23: A proposal for *Bacillus subtilis* subsp. *subtilis* subsp. nov, and *Bacillus subtilis* subsp. *spizizenii* subsp. nov. *International Journal of Systematic Bacteriology* **49**: 1211–1215. 10.1099/00207713-49-3-1211

Nihorimbere, V., Cawoy, H., Seyer, A., Brunelle, A., Thonart, P. & Ongena, M. (2012) Impact of rhizosphere factors on cyclic lipopeptide signature from the plant beneficial strain *Bacillus amyloliquefaciens* S499. *FEMS Microbiology Ecology* **79**: 1, 176–191. 10.1111/j.1574-6941.2011.01208.x

Okada, M. (2011) Post-translational isoprenylation of tryptophan. *Bioscience, Biotechnology, and Biochemistry* **75**: 8, 1413–1417. 10.1271/bbb.110087

Okada, M., Sugita, T. & Abe, I. (2017) Posttranslational isoprenylation of tryptophan in bacteria. *Beilstein Journal of Organic Chemistry* **13**: 338–346. 10.3762/bjoc.13.37

Okada, M., Nakamura, Y., Hayashi, S., Ozaki, K. & Usami, S. (2015) Chemical structure and biological activity of a quorum sensing pheromone from *Bacillus subtilis* subsp. *natto*. *Bioorganic and Medicinal Chemistry Letters* **25**: 19, 4293–4296. http://doi.org/10.1016/j.bmcl.2015.07.083

Okada, M., Yamaguchi, H., Sato, I., Tsuji, F., Dubnau, D. & Sakagami, Y. (2008) Chemical structure of posttranslational modification with a farnesyl group on tryptophan. *Bioscience, Biotechnology, and Biochemistry* **72**: 3, 914–918. 10.1271/bbb.80006

Okada, M., Sato, I., Cho, S. J., Iwata, H., Nishio, T., Dubnau, D. & Sakagami, Y. (2005) Structure of the *Bacillus subtilis* quorum-sensing peptide pheromone ComX. *Nature Chemical Biology* **1**: 1, 23–24. 10.1038/nchembio709

Oslizlo, A., Stefanic, P., Dogsa, I. & Mandić-Mulec, I. (2014) Private link between signal and response in *Bacillus subtilis* quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 4, 1586–1591. 10.1073/pnas.1316283111

Oslizlo, A., Stefanic, P., Vatovec, S., Beigot Glaser, S., Rupnik, M. & Mandić-Mulec, I. (2015) Exploring ComQXPA quorum-sensing diversity and biocontrol potential of *Bacillus* spp. isolates from tomato rhizoplane. *Microbial Biotechnology* **8**: 3, 527–540. 10.1111/1751-7915.12258

Parashar, V., Konkol, M. A., Kearns, D. B. & Neiditch, M. B. (2013) A plasmid-encoded phosphatase regulates *Bacillus subtilis* biofilm architecture, sporulation, and genetic competence. *Journal of Bacteriology* **195**: 10, 2437–2448. 10.1128/jb.02030-12

Piazza, F., Tortosa, P. & Dubnau, D. (1999) Mutational analysis and membrane topology of ComP, a quorum-sensing histidine kinase of *Bacillus subtilis* controlling competence development. *Journal of Bacteriology* **181**: 15, 4540–4548. 10.1128/JB.181.15.4540-4548.1999

Pottathil, M. & Lazazzera, B. A. (2003) The extracellular Phr peptide-Rap phosphatase signaling circuit of *Bacillus subtilis*. *Frontiers in Bioscience* **8**: d32–45. 10.2741/913

Pottathil, M., Jung, A. & Lazazzera, B. A. (2008) CSF, a species-specific extracellular signaling peptide for communication among strains of *Bacillus subtilis* and *Bacillus mojavensis*. *Journal of Bacteriology* **190**: 11, 4095–4099. 10.1128/jb.00187-08

Ramey, B. E., Koutsoudis, M., von Bodman, S. B. & Fuqua, C. (2004) Biofilm formation in plant-microbe associations. *Current Opinion in Microbiology* **7**: 602–609. https://doi.org/10.1016/j.mib.2004.10.014

Reva, O. N., Dixelius, C., Meijer, J. & Priest, F. G. (2004) Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *FEMS Microbiology Ecology* **48**: 249–259. 10.1016/j.femsec.2004.02.003

Reva, O. N., Swanevelder, D. Z. H., Mwita, L. A., Mwakilili, A. D., Muzondiwa, D., Joubert, M., Chan, W. Y., Lutz, S., Ahrens, C. H., Avdeeva, L. V., Kharkhota, M. A., Tibuhwa, D., Lyantagaye, S., Vater, J., Borriss, R. & Meijer, J. (2019) Genetic, epigenetic and phenotypic diversity of four *Bacillus velezensis* strains used for plant protection or as probiotics. *Frontiers in Microbiology* **10**: 2610, 1–25. 10.3389/fmicb.2019.02610

Richter, M. & Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 45, 19126–19131. 10.1073/pnas.0906412106

Roberts, M. S. & Cohan, F. M. (1995) Recombination and migration mates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* **49**: 6, 1081–1094. https://doi.org/10.1111/j.1558-5646.1995.tb04435.x

Roggiani, M. & Dubnau, D. (1993) ComA, a phosphorylated response regulator protein of *Bacillus subtilis*, binds to the promoter region of *srf*A. *Journal of Bacteriology* **175**: 10, 3182–3187. 10.1128/jb.175.10.3182-3187.1993

Romero, D., Aguilar, C., Losick, R. & Kolter, R. (2010) Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5, 2230–2234. 10.1073/pnas.0910560107

Romero, D., Vlamakis, H., Losick, R. & Kolter, R. (2011) An accessory protein required for anchoring and assembly of amyloid fibres in *B. subtilis* biofilms. *Molecular Microbiology* **80**: 5, 1155–1168. 10.1111/j.1365-2958.2011.07653.x

Rooney, A. P., Price, N. P. J., Ehrhardt, C., Swezey, J. L. & Bannan, J. D. (2009) Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology* **59**: 2429–2436. 10.1099/ijs.0.009126-0

Rudrappa, T., Biedrzycki, M. L. & Bais, H. P. (2008) Causes and consequences of plant-associated biofilms. *FEMS Microbiology Ecology* **64**: 2, 153–166. 10.1111/j.1574-6941.2008.00465.x

Ruiz-Garcia, C., Bejar, V., Martinez-Checa, F., Llamas, I. & Quesada, E. (2005) *Bacillus velezensis* sp. nov., a surfactant-producing bacterium isolated from the river Velez in Malaga, southern Spain. *International Journal of Systematic and Evolutionary Microbiology* **55**: 1, 191–195. 10.1099/ijs.0.63310-0

Schauder, S. & Bassler, B. L. (2001) The languages of bacteria. *Genes & Development* **15**: 12, 1468–1480. 10.1101/gad.899601

Schleifer, K.-H. (2009b) The Firmicutes. *Bergey's Manual of Systematic Bacteriology* Vol. 3 (Parte, A. C., Whitman, W. B., De Vos, P., Garrity, G. M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A. & Schleifer, K.-H., eds.), pp. 19–128. Springer, New York, NY. 10.1002/9781118960608

Schleifer, K. H. (2009a) Classification of Bacteria and Archaea: Past, present and future. *Systematic and Applied Microbiology* **32**: 8, 533–542. https://doi.org/10.1016/j.syapm.2009.09.002

Schultz, D., Wolynes, P. G., Ben Jacob, E. & Onuchic, J. N. (2009) Deciding fate in adverse times: Sporulation and competence in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 50, 21027–21034. 10.1073/pnas.0912185106

Shi, X. & Zhu, X. (2009) Biofilm formation and food safety in food industries. *Trends in Food Science & Technology* **20**: 9, 407–413. https://doi.org/10.1016/j.tifs.2009.01.054

Sikorski, J. & Nevo, E. (2005) Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 44, 15924–15929. 10.1073/pnas.0507944102

Smith, P. & Schuster, M. (2019) Public goods and cheating in microbes. *Current Biology* **29**: 11, R442–R447. https://doi.org/10.1016/j.cub.2019.03.001

Solomon, J. M. & Grossman, A. D. (1996) Who's competent and when: Regulation of natural genetic competence in bacteria. *Trends in Genetics* **12**: 4, 150–155. https://doi.org/10.1016/0168-9525(96)10014-7

Solomon, J. M., Lazazzera, B. A. & Grossman, A. D. (1996) Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes & Development* **10**: 16, 2014–2024. 10.1101/gad.10.16.2014

Solomon, J. M., Magnuson, R., Srivastava, A. & Grossman, A. D. (1995) Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes & Development* **9**: 5, 547–558. 10.1101/gad.9.5.547

Stackebrandt, J. E. (2006) Taxonomic parameters revisited: Tarnished gold standards *Microbiology Today* 33, 152–155.

Stefanic, P. & Mandić-Mulec, I. (2009) Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. *Journal of Bacteriology* **191**: 6, 1756–1764. 10.1128/jb.01290-08

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F. M. & Mandić-Mulec, I. (2012) The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. *Environmental Microbiology* **14**: 6, 1378–1389. 10.1111/j.1462-2920.2012.02717.x

Sturme, M. H., Kleerebezem, M., Nakayama, J., Akkermans, A. D., Vaugha, E. E. & de Vos, W. M. (2002) Cell to cell communication by autoinducing peptides in Gram-positive bacteria. *Antonie Van Leeuwenhoek* **81**: 1–4, 233–243. 10.1023/A:1020522919555

Swiecicka, I., Bartoszewicz, M., Kasulyte-Creasey, D., Drewnowska, J. M., Murawska, E., Yernazarova, A., Lukaszuk, E. & Mahillon, J. (2013) Diversity of thermal ecotypes and

potential pathotypes of *Bacillus thuringiensis* soil isolates. *FEMS Microbiology Ecology* **85**: 2, 262–272. 10.1111/1574-6941.12116

Teeling, H., Meyerdierks, A., Bauer, M., Rudolf, A. & Glöckner, F. O. (2004) Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environmental Microbiology* **6**: 9, 938–947. doi:10.1111/j.1462-2920.2004.00624.x

Tortosa, P. & Dubnau, D. (1999) Competence for transformation: A matter of taste. *Current Opinion in Microbiology* **2**: 6, 588–592. https://doi.org/10.1016/S1369-5274(99)00026-0

Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandić-Mulec, I. & Dubnau, D. (2001) Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. *Journal of Bacteriology* **183**: 2, 451–460. 10.1128/JB.183.2.451-460.2001

Tran, L. S., Nagai, T. & Itoh, Y. (2000) Divergent structure of the ComQXPA quorum-sensing components: Molecular basis of strain-specific communication mechanism in *Bacillus subtilis*. *Molecular Microbiology* **37**: 5, 1159–1171. 10.1046/j.1365-2958.2000.02069.x

Velmourougane, K., Prasanna, R. & Saxena, A. K. (2017) Agriculturally important microbial biofilms: Present status and future prospects. *Journal of Basic Microbiology* **57**: 7, 548–573. 10.1002/jobm.201700046

Vlamakis, H., Aguilar, C., Losick, R. & Kolter, R. (2008) Control of cell fate by the formation of an architecturally complex bacterial community. *Genes & Development* **22**: 7, 945–953. 10.1101/gad.1645008

Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. (2013) Sticking together: Building a biofilm the *Bacillus subtilis* way. *Nature Reviews Microbiology* **11**: 3, 157–168. 10.1038/nrmicro2960

Ward, B. B. (2002) How many species of prokaryotes are there? *Proceedings of the National Academy of Sciences of the United States of America* **99**: 16, 10234–10236. 10.1073/pnas.162359199

Ward, D. M. (2006) A macrobiological perspective on microbial species. Vol. 1 pp. 269–278. *Microbe Magazine*. American Society for Microbiology. 10.1128/microbe.1.269.1

Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kuhl, M. & Grossman, A. (2007) Genomics, environmental genomics and the issue of microbial species. *Heredity* **100**: 2, 207–219. 10.1038/sj.hdy.6801011

Waters, C. M. & Bassler, B. L. (2005) Quorum sensing: Cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology* **21**: 319–346. 10.1146/annurev.cellbio.21.012704.131001

Weinrauch, Y., Penchev, R., Dubnau, E., Smith, I. & Dubnau, D. (1990) A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-transduction systems. *Genes & Development* **4**: 5, 860–872. 10.1101/gad.4.5.860

Williams, P. (2007) Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* **153**: 3923–3938. 10.1099/mic.0.2007/012856-0

Williams, P., Winzer, K., Chan, W. C. & Camara, M. (2007) Look who's talking: Communication and quorum sensing in the bacterial world. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **362**: 1483, 1119–1134. 10.1098/rstb.2007.2039

Yaryura, P. M., Leon, M., Correa, O. S., Kerber, N. L., Pucheu, N. L. & Garcia, A. F. (2008) Assessment of the role of chemotaxis and biofilm formation as requirements for colonization of roots and seeds of soybean plants by *Bacillus amyloliquefaciens* BNM339. *Current Microbiology* **56**: 6, 625–632. 10.1007/s00284-008-9137-5

Yu, Y., Yan, F., Chen, Y., Jin, C., Guo, J. H. & Chai, Y. (2016) Poly-gamma-glutamic acids contribute to biofilm formation and plant root colonization in selected environmental isolates of *Bacillus subtilis*. *Frontiers in Microbiology* **7**: 1811, 1–15. 10.3389/fmicb.2016.01811

Zhang, N., Yang, D., Kendall, J. R. A., Borriss, R., Druzhinina, I. S., Kubicek, C. P., Shen, Q. & Zhang, R. (2016) Comparative genomic analysis of *Bacillus amyloliquefaciens* and *Bacillus subtilis* reveals evolutional traits for adaptation to plant-associated habitats. *Frontiers in Microbiology* **7**: 2039, 1–14. 10.3389/fmicb.2016.02039

CHAPTER TWO

IN SILICO EVALUATION OF MLSA AND *COM*QXP PCR-RFLP TO DISTINGUISH PLANT-ASSOCIATED *BACILLUS SUBTILIS SENSU LATO*

2.1. Introduction

Bacterial species and populations demonstrate genetic and functional variances based on their ecology and geography (Ward *et al.*, 2007). Subpopulation variability within a species has given rise to the "ecological species concept", which has implications for ecology and niche specialisation and how these are accommodated with regards to speciation and taxonomic conventions (Cohan, 2005, Koeppel *et al.*, 2008). The consequences of subpopulation divergence on bacterial taxonomy is still being debated across many microbial species; the genus *Bacillus* being one such example (Kopac *et al.*, 2014). The ubiquitous genus *Bacillus* comprises phenotypically heterogeneous species occupying a range of environmental niches. These species, in spite of highly similar 16S rRNA gene sequences, display a varied and complicated taxonomy (Fritze, 2004, Earl *et al.*, 2008, Connor *et al.*, 2010, Fajardo-Cavazos *et al.*, 2014, Mandić-Mulec *et al.*, 2015). The housekeeping gene heterogeneity and functional diversity demonstrated by *Bacillus* spp. has led to the description of *Bacillus* species complexes, or *sensu lato* groupings, to accommodate these closely-related groups of taxa (Fritze, 2004).

Within these broad groupings of related taxa there exists evidence for further diversification into niche-adapted ecologically distinct groupings (i.e. ecotypes), most notably within members of the *B. subtilis sensu lato* (Reva *et al.*, 2004, Alcaraz *et al.*, 2010, Kopac *et al.*, 2014, Mandić-Mulec *et al.*, 2015). The taxonomic implications of ecological variation amongst members of the *B. subtilis sensu lato* has seen this group significantly expanded and revised over the last decade (Fan *et al.*, 2017). One such aspect surrounds the taxonomic debates of plant-associated *B. amyloliquefaciens* subsp. *plantarum* (Dunlap *et al.*, 2015, Dunlap *et al.*, 2016, Fan *et al.*, 2019). Adaptation to a plant-associated lifestyle and capabilities in plant health promotion have seen the *B. amyloliquefaciens* subsp. *plantarum* elevated to species level as *B.*

velezensis, with the *B. velezensis* strain FZB42 recently declared to be the model strain for plant-associated Gram-positive bacteria (Chen *et al.*, 2009, Fan *et al.*, 2019).

Ecological functional divergence within the *B. subtilis sensu lato* has also been associated with quorum-sensing diversification (Stefanic *et al.*, 2012). One of the pathways *Bacillus* spp. employs quorum sensing involves the ComQXPA quorum-sensing cassette, which demonstrates hypervariability between populations and forms the basis for distinct communication groupings known as pherotypes (Ansaldi *et al.*, 2002). The relationship between ecotypes and pherotypes has been explored in several species of *B. subtilis sensu lato*, primarily in *B. subtilis* and *B. mojavensis* (Stefanic and Mandić-Mulec, 2009, Stefanic *et al.*, 2012, Oslizlo *et al.*, 2015). These previous studies of pherotype specialisation and ecological adaptation amongst soil-dwelling and plant-associated *Bacillus* spp. provides grounds to propose a link between variability in ComX pheromones and niche adaptation in plant-associated *B. subtilis.* Pherotype variation in plant-associated *B. velezensis* has not been investigated, despite the value of these beneficial bacteria in plant health, and the interplay of ecotype and pherotype in plant-adapted *B. velezensis* presents an enticing prospect for study.

The present study aimed to apply existing GenBank genomic data and in silico approaches to investigate subspecies genomic variation and pherotype diversity in selected *B. velezensis* and *B. subtilis* reference strains, with an emphasis on plant-associated representatives. Subpopulation diversification within representative *Bacillus subtilis sensu lato* member genomes was carried out in silico using Multilocus Sequence Analysis (MLSA) of housekeeping gene sequence data available in the GenBank database. Pherotype variation amongst these representatives was assessed indirectly by examining the DNA sequences of the quorum sensing *comQXPA* gene loci. From these data, and by comparison to existing *Bacillus* spp. pherotypes, a high-throughput pherotyping PCR-RFLP protocol was developed in silico to identify and group putative pherotypes amongst the *Bacillus* strains studied.

2.2. Materials and Methods

2.2.1. Gene sequences for *Bacillus subtilis sensu lato* group representatives

Gene sequences for the MLSA and *comQXP* pherotyping experiments study were obtained from the National Center for Biotechnology Information (NCBI) GenBank database (URL: https://www.ncbi.nlm.nih.gov/genbank/) (Last Accessed: 30 September 2020). For the purposes of this study, *B. amyloliquefaciens* subsp. *plantarum* strains were considered as later heterosynonyms of *B. velezensis*, acknowledging current taxonomic convention (Dunlap *et al.*, 2016). The synonymity of *B. amyloliquefaciens* subsp. *plantarum* and *B. velezensis* is relatively recent (Fan *et al.*, 2019), and the former naming convention persists in some instances in the GenBank database. For the purposes of this research, the term "*B. velezensis*" is intended to also encompass heterotypic synonyms *B. amyloliquefaciens* subsp. *plantarum* and *B. methylotrophicus*. Gene sequences for this study were obtained from the whole genome record of strains selected a**s** genomes representative of the *Bacillus subtilis sensu lato* grouping. An emphasis was placed upon the inclusion of plant-associated strains and, in particular, included a range of *B. velezensis* strains. Also included were strains from other environmental sources and well-known laboratory strains (Table 2.1), and with *B. cereus* ATCC 14579 included as a taxonomic outgroup**.** Pherotyping experiments used the *comQXP* sequence data of these strains and also included the partial quorum-sensing loci sequences of known *Bacillus* spp. reference pherotype strains (Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002) (Table 2.2.).

Table 2.1. Representative genomes of *Bacillus subtilis sensu lato* selected for this study and sourced from the NCBI GenBank database.

Species and Strain	GenBank Accession No.	Origin / Source
B. velezensis SQR9	CP006890.1	Cucumber rhizosphere soil
B. amyloliquefaciens S499	CP014700.1	Plant soil
B. velezensis L-S60	CP011278.1	Soil
B. velezensis GYL4	NZ_CP020874.1	Pepper plant
B. velezensis NJAU-Z9	NZ_CP022556.1	Field soil
B. velezensis M75	NZ_CP016395.1	Mushroom cultivation substrate
B. velezensis MH25	CP034176.1	Rhizosphere soil
B. velezensis BCSo1	CP034037.1	*Danio rerio*
B. velezensis 1B-23	CP033967.1	Potato rhizosphere
B. velezensis ANSB01E	CP036518.1	Chicken intestine
B. velezensis ZF2	CP032154.1	Cucumber plants
B. velezensis JT3-1	NZ_CP032506.1	Yak faeces
B. amyloliquefaciens IT-45	CP004065.1	Commercial horticulture strain
B. amyloliquefaciens XH7	CP002927.1	Purine nucleoside production
B. amyloliquefaciens subsp. *amyloliquefaciens* DSM 7	NC_014551.1	DSMZ[¥] type strain
B. amyloliquefaciens LFB112	NC_023073	Chinese herbs
B. amyloliquefaciens Y2	CP003332.1	Wheat rhizosphere
B. amyloliquefaciens subsp. *plantarum* CAU-B946	NC_016784.1	Rice rhizosphere
B. velezensis FZB42	NC_009725.2	Beet field soil
B. amyloliquefaciens subsp. *plantarum* UCMB5113	NC_022081.1	Soil
B. amyloliquefaciens subsp. *plantarum* UCMB5036	NC_020410.1	*Gossypium barbadense*
B. amyloliquefaciens subsp. *plantarum* YAU B9601-Y2	NC_017061.1	Wheat rhizosphere
Bacillus sp. JS	NC_017743.1	*Miscanthus* soil
B. subtilis QB928	NC_018520.1	Genome reference
B. subtilis subsp. *spizizenii* TU-B-10	NC_016047.1	Tunisian desert
B. subtilis subsp. *spizizenii* W23	CP002183.1	BGSC* reference strain
B. subtilis subsp. *subtilis* 168	NC_000964.3	Laboratory reference strain
B. subtilis subsp. *subtilis* NCIB 3610	CP020102.1	Type strain
B. cereus ATCC 14579	NC_004722.1	ATCC[#] type strain

[¥] Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany. * Bacillus Genetic Stock Center, The Ohio State University, Columbus, Ohio, U. S. A. [#] American Type Culture Collection, Manassas, Virginia, U .S. A.

Table 2.2. Partial quorum-sensing gene loci sequences of *Bacillus subtilis sensu lato* pherotypes selected for this study and sourced from the NCBI GenBank database.

Species and Strain*	GenBank Accession No.
B. subtilis RO-FF-1	AF456130.1
B. subtilis RO-PP-2	AF456131.1
B. subtilis RO-A-4	AF456132.1
B. subtilis RO-DD-2	AF456133.1
B. mojavensis RO-C-2	AF456134.1
B. mojavensis RO-B-2	AF456135.1
B. subtilis RS-D-2	AF456136.1
B. subtilis RO-F-3	AF456137.1
B. subtilis RO-E-2	AF456138.1
B. mojavensis RO-H-1	AY003901.1
B. subtilis RS-B-1	AY003900.1
B. subtilis BEST195 (~NAF4)[#]	NC_017196.2

***** Pherotypes in *Bacillus subtilis sensu lato* as described by Tortosa *et al.*, 2001 and Ansaldi *et al.*, 2002. [#]Applied as counterpart to *B. subtilis* NAF4 pherotype (Nishito *et al*., 2010) owing to lack of NAF4 availability in the GenBank database.

2.2.2. Using multilocus analysis to differentiate *Bacillus subtilis sensu lato*

An MLSA of pertinent plant-associated strains of the *B. subtilis sensu lato* group was conducted using housekeeping genes from existing whole-genome sequences (Table 2.1.). Nine housekeeping gene sequence candidates were selected for this analysis based on genes used in previous studies to distinguish members of this taxonomic grouping (Rooney *et al.*, 2009, Stefanic and Mandić-Mulec, 2009, Borriss, 2011). These genes were 16S rRNA subunit (16S rRNA), gyrase subunits alpha (*gyrA*) and beta (*gyrB*), RNA polymerase subunit beta (*rpoB*), phosphoribosylaminoimidazolecarboxamide formyltransferase (*purH*), DNA polymerase III subunit alpha (*polC),* 60 kDa heat-shock protein groEL (*groEL*), DNA chaperone protein (*dnaJ*), and chemotaxis protein histidine kinase (*cheA*).

The MLSA analysis of the housekeeping gene sequences and selected reference strains was conducted using MEGA X software (version 10.0.5.) (Pennsylvania State University) (Kumar *et al.*, 2018). Alignments and phylogenies were generated for the housekeeping gene sequences using MEGA X software and aligned in MEGA X using MUSCLE (Edgar, 2004). After individual housekeeping gene sequence alignment, all the sequences were concatenated using MEGA X and phylogeny generation was subsequently performed. Phylogenies were generated using the Maximum Likelihood method and Kimura 2-parameter model (Kimura, 1980) and phylogeny testing applied a bootstrap method of 500 replicates. The initial trees applied to the heuristic search were obtained by Neighbour-Join and BioNJ algorithms, which were applied to a pairwise distances matrix. This matrix was estimated using the Maximum Composite Likelihood approach, and then selecting the topology with a higher log likelihood value. Individual gene sequence phylogenies were also generated under the above-mentioned conditions and parameters for comparative purposes. A smaller-scale MLSA was also generated using only the 16S rRNA, *gyrA*, *rpoB*, and *dnaJ* sequence data.

2.2.3. In silico pherotyping of *Bacillus subtilis sensu lato* using *comQXP* sequences

Pherotyping of strains was explored using gene sequence data of the *comQXP* quorum-sensing gene region from the representative whole genome *Bacillus* strains used for MLSA generation (Table 2.1.) and included *comQ* gene sequence data entries for known *Bacillus* spp. pherotypes from the NCBI GenBank database (Ansaldi *et al.*, 2002). A rapid-pherotyping technique was also developed using simulated PCR and PCR-RFLP of the *comQXP* gene region for genus *Bacillus* representatives using an online in silico PCR tool (URL: http://insilico.ehu.es/PCR) (Bikandi *et al.*, 2004, San Millán *et al.*, 2013). The PCR simulation applied primers for the *comQXP* region, which have been published for *B. subtilis* (Stefanic and Mandić-Mulec, 2009); whereas, *B. amyloliquefaciens* and *B. velezensis comQXP* sequence variations necessitated that a PCR primer set be designed for these species.

2.2.3.1. comQXP *gene region PCR primers*

The two *Bacillus* species groups being investigated (viz., *B. subtilis* and *B. velezensis*) required different primer pairs to amplify their respective *comQXP* gene regions. The primers for *B. subtilis* utilised an existing primer pair developed for this species, the forward primer

UniComQ1 5'-GGGAGGGGGGAAGTCGTTATTG-3' and reverse primer P1 5'-AAGAACCGAATCGTGGAGATCGCG-3' (Stefanic and Mandić-Mulec, 2009). A novel *comQXP* primer set was derived for the *B. velezensis* species using the online tool PrimerBLAST (URL: https://www.ncbi.nlm.nih.gov/tools/primer-blast) (Ye *et al.*, 2012) applying reference genomes *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7 (NC_014551.1) and *B. velezensis* FZB42 (NC_009725.2). The *comQXP* primer pair candidates were evaluated using the online in silico PCR tool (Bikandi *et al.*, 2004, San Millán *et al.*, 2013) for coverage of all *B. amyloliquefaciens* representatives available on the tool (which included *B. velezensis* strains). The following primer pair was chosen: forward primer ComBamyF 5'-GACATCTGCTCATTCATCTGATGAAG-3' and reverse primer ComBamyR 5'-CAGAGAAACAACCGACTCATTACG-3'.

2.2.3.2. In silico PCR-RFLP of the com*QXP gene region*

A simulated rapid-pherotyping PCR-RFLP technique was developed using the *comQXP* gene region for the genus *Bacillus* representatives available in an online in silico PCR tool (URL: http://insilico.ehu.es/PCR) (Bikandi *et al.*, 2004, San Millán *et al.*, 2013). The simulated PCR parameters evaluated both aforementioned *comQXP* primer sets against all strains of the *Bacillus* genus represented in the tool, allowing for an amplicon of 4 kb, and no mismatches allowed. Restriction digests of these two sets of amplicons were then simulated for each primer set using the in silico PCR-RFLP tool. The restriction enzymes selected for the simulated PCR-RFLP of the *comQXP* gene loci were chosen empirically by testing them against the amplicons generated for the *Bacillus* genus representatives included in the tool.

Phylogenetic analysis was then undertaken of the projected amplicons (as downloaded from the in silico tool output) and for the *comQXP* region for selected *Bacillus* spp. reference genomes. The gene sequences were aligned using MUSCLE in MEGA X (version 10.0.5.) (Pennsylvania State University) (Kumar *et al.*, 2018). Sequence lengths were homogenised using BioEdit (Hall, 1999), and phylogenetic analysis was then conducted in MEGA X using the parameters specified in Section 2.2.2. The *comQXP* phylogeny comprised simulated amplicons from the in silico PCR tool along with *comQXP* sequences of the MLSA strains

(Table 2.1.) obtained from GenBank. A comparative *comQ* phylogeny was generated which included existing pherotype *comQ* sequence data (Table 2.2.), also obtained from GenBank.

The suitability of simulated PCR-RFLP outputs were evaluated according to various parameters, which included: enzyme representation between species, number of bands per profile, and overall fingerprint distinguishability and readability. Strain-level discrimination achievable by the restriction digests was evaluated by two criteria: (1) capacity to distinguish *Bacillus* sp. JS from the remainder of the *B. subtilis* strains and (2) discrimination of *B. amyloliquefaciens* subsp. *plantarum* (*B. velezensis*) FZB42 from the remainder of *B. amyloliquefaciens* strains. These data allowed the selection of a suite of restriction enzyme candidates for PCR-RFLP of the *comQXP* region for rapid pherotyping purposes.

2.3. Results

2.3.1. Multilocus analysis of *Bacillus subtilis sensu lato*

As expected, the 16S rRNA gene phylogeny (Figure 2.1.) demonstrated limited ability to resolve between the closely-related members of the *B. subtilis sensu lato* grouping chosen for this experiment. The phylogenetic tree showed a distinct separation of clades between the two species *B. amyloliquefaciens* and *B. subtilis* and their close relatives. However, in the clustering of the *B. amyloliquefaciens* representatives, the *B. velezensis* strains were largely indistinguishable from the *B. amyloliquefaciens* species. It is unclear whether this may be attributed to older naming conventions of strains which may be reclassified as *B. velezensis*; e.g. *B. amyloliquefaciens* S499 grouping with *B. velezensis* SQR9 and not with the XH7 and DSM 7 nomenclature counterparts.

The MLSA clustering (Figure 2.2.A) comprising nine housekeeping genes was able to resolve polymorphisms in the gene sequence data to subspecies level and afforded greater subgrouping resolution when compared to the clustering observed for the conventional 16S rRNA gene sequence phylogeny (Figure 2.1.). The MLSA phylogeny correctly grouped the closely-related

strains of *B. subtilis* QB928, *B. subtilis* subsp. *subtilis* 168, and *B. subtilis* subsp. *subtilis* NCIB 3610 (Dedonder *et al.*, 1977, Yu *et al.*, 2012). Gene polymorphisms distinguishing the plant-associated *Bacillus* sp. JS was also evident, and placed this strain between the clusters related to *B. subtilis* subsp. *spizizenii* representatives and the *B. subtilis* strains. The *B. velezensis* representatives were not all clustered together, but rather formed several groups within the larger *B. amyloliquefaciens* clade, with a definite a separation of *Bacillus amyloliquefaciens* strains (DSM 7 and XH7) from the *B. velezensis* representatives. Amongst the diverse representative selection of *B. velezensis*, the origin of the strains did not have any impact on clustering as the clades contain strains from environmental and plant-associated sources. For example; the clade comprising *B. velezensis* ZF2, JT3-1, ANSB01E, SQR9 and GYL4 comprise origins of cucumber, yak faeces, chicken intestine, cucumber plant and pepper plant, respectively.

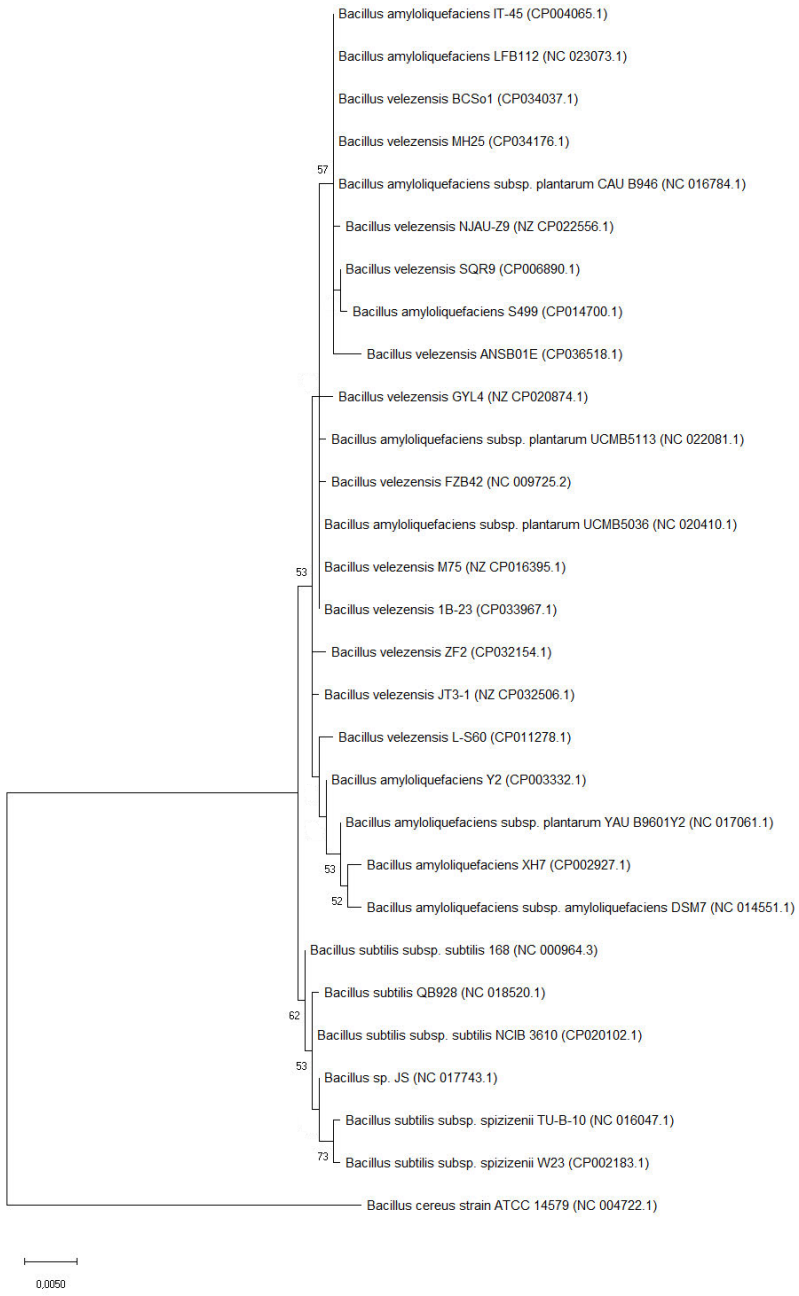
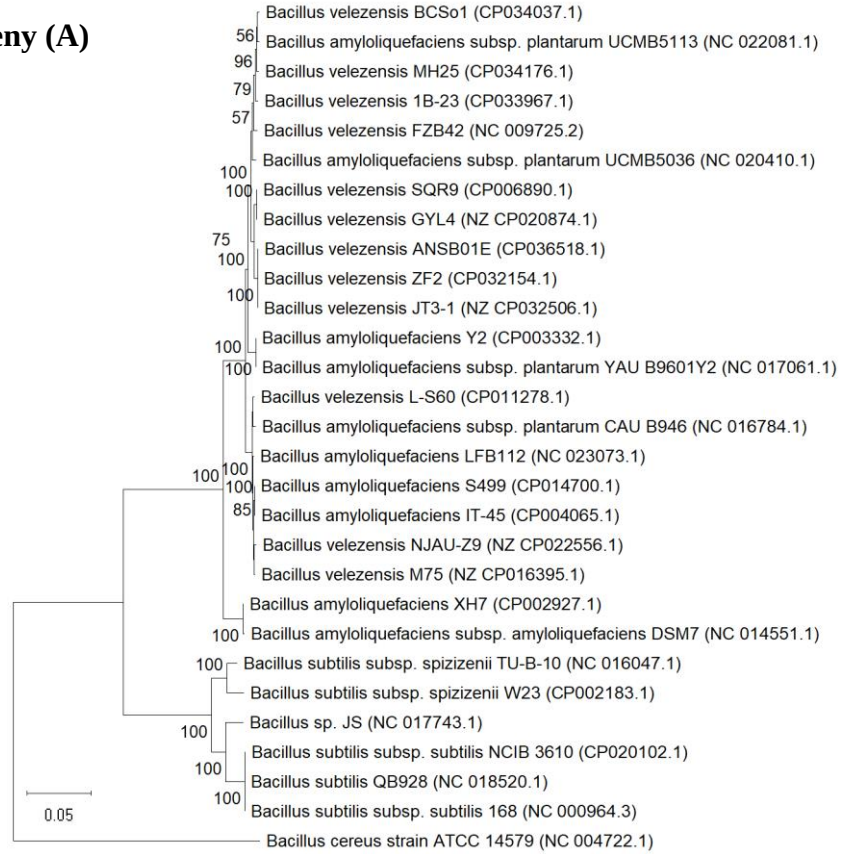


Figure 2.1. Phylogeny of 16S rRNA gene sequences for selected strains of *Bacillus subtilis sensu lato.* The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-2820.99) as shown. The numbering next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 1506 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

Phylogeny (A)



Phylogeny (B)

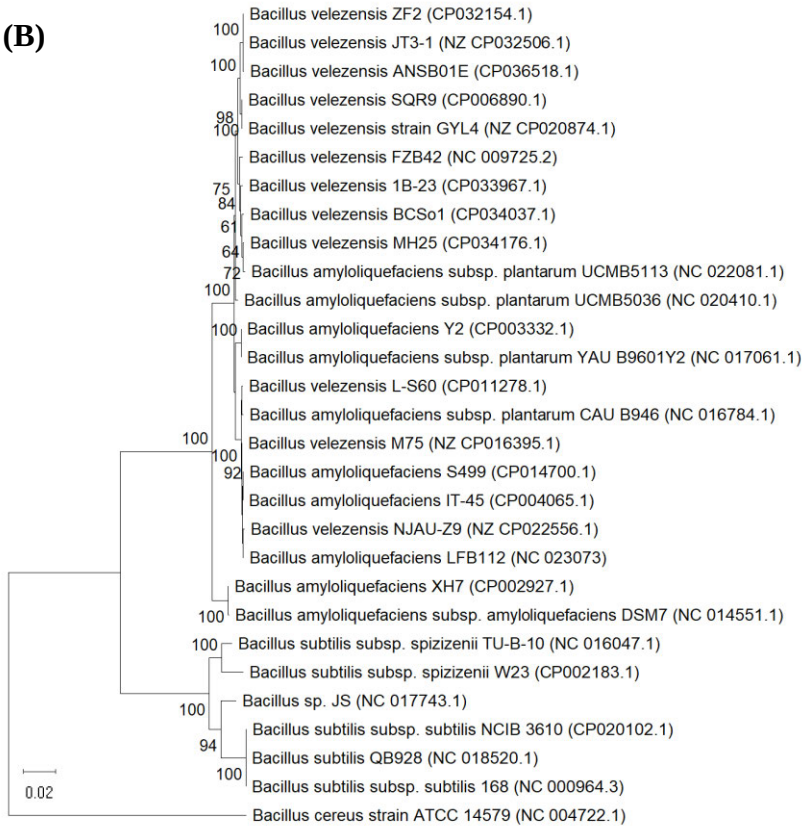


Figure 2.2. MLSA phylogenies comprised of housekeeping gene sequences for selected strains of *Bacillus subtilis sensu lato* comparing the outputs for: (A) Nine gene sequences (viz., 16S rRNA-*cheA*-*dnaJ*-*groEL*-*gyrA*-*gyrB*-*polC*-*purH*-*rpoB*) and (B) Four gene sequences (viz., 16S rRNA-*dnaJ*-*gyrA*-*rpoB*) respectively. The evolutionary analysis was conducted in MEGA X using the Maximum Likelihood method and Kimura 2-parameter model. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together. Phylogeny (B) applied 29 nucleotide sequences with 19917 positions in the final dataset, the tree shown represents the highest log likelihood (-74892.52). Phylogeny (B) applied 29 nucleotide sequences with 8357 positions in the final dataset, the tree represents the highest log likelihood (-27992.53).

Individual housekeeping gene phylogenies were generated to compare subpopulation-level resolution and clustering of the chosen *Bacillus* spp. strains (See Appendix A). Permutations of housekeeping gene combinations were also evaluated to determine which gene combination best represented the clustering trends observed in the full nine-gene MLSA (Data not shown). The individual gene phylogenetic trees showed minor variations in clustering and confidences but, broadly speaking, all of these genes reflect the MLSA the clusters observed. Four housekeeping genes (viz., 16S rRNA, *gyrA*, *rpoB*, and *dnaJ*) were selected for a smaller-scale MLSA (Figure 2.2.B) on the basis of discrimination capability between *B. amyloliquefaciens* and *B. velezensis*, and between *B. subtilis* and closely-related subspecies groups. The abridged MLSA retained the sensitivity to variations within the *B. velezensis* representatives, and clustered *B. amyloliquefaciens* representatives separately to the remaining *B. velezensis* representatives. Also reproduced was the three-clade differentiation amongst those strains most closely-related to *B. subtilis sensu stricto*: (1) *B. subtilis* subsp. *spizizenii* TU-B-10 and W23; (2) *B. subtilis* QB928, *B. subtilis* subsp. *subtilis* NCIB3610 and 168, and the placement of (3) *Bacillus* sp. JS between *B. subtilis* and *B. subtilis* subsp. *spizizenii*.

2.3.2. Pherotyping in silico using *comQXP* gene sequences and PCR-RFLP

ComX divergence amongst strains of *Bacillus* spp. may be indirectly assessed by targeting gene sequence data for the *comQXP* region, which encodes the immature ComX pheromone, and its associated machinery. The *comQXP* sequence phylogenies included sequence data for the representative genomes applied in the MLSA (Figure 2.3.), comparison of *comQ* with known pherotypes (Figure 2.4.) and in silico simulated *comQXP* amplicons (Figure 2.5.), and indicated that diversification in communication gene loci was not strictly species-limited. Representative clusters of each *B. subtilis, B. velezensis,* and *B. amyloliquefaciens* species were observed within the various clades, which suggests divergence in these quorum sensing sequences that surpasses species delineation. A phylogeny generated using the simulated amplicons from the in silico PCR tool output and the corresponding gene regions obtained from the reference genomes in GenBank (data not shown) indicated that the in silico tool output and regions of primer targets were accurate and representative.

The *comQ* phylogeny (Figure 2.4.) also alludes to greater diversity within communication genes amongst environmental *Bacillus* than just those previously-described pherotypes (Ansaldi *et al.*, 2002). These "known" pherotypes were observed to group largely amongst themselves, and there are only a handful of instances where they cluster closely to any of the representative strains: (A) *B. subtilis* BSn5 with pherotypes RS-D-2 and RO-F-3 with *B. subtilis* subsp. *spizizenii* TU-B-10; and (B) *Bacillus* sp. JS with pherotype RO-B-2. This phylogeny suggests that there is significant ComX diversity in the *B. amyloliquefaciens* and *B. velezensis comQ* than those described pherotypes in the *B. subtilis* and *B. mojavensis* species. Furthermore, this diversity suggests that the ComX pheromones employed by *B. amyloliquefaciens* and *B. velezensis* are novel and distinct from the known pherotypes, as none of the ten clusters of *B. amyloliquefaciens* or *B. velezensis* strains clades include a known pherotype.

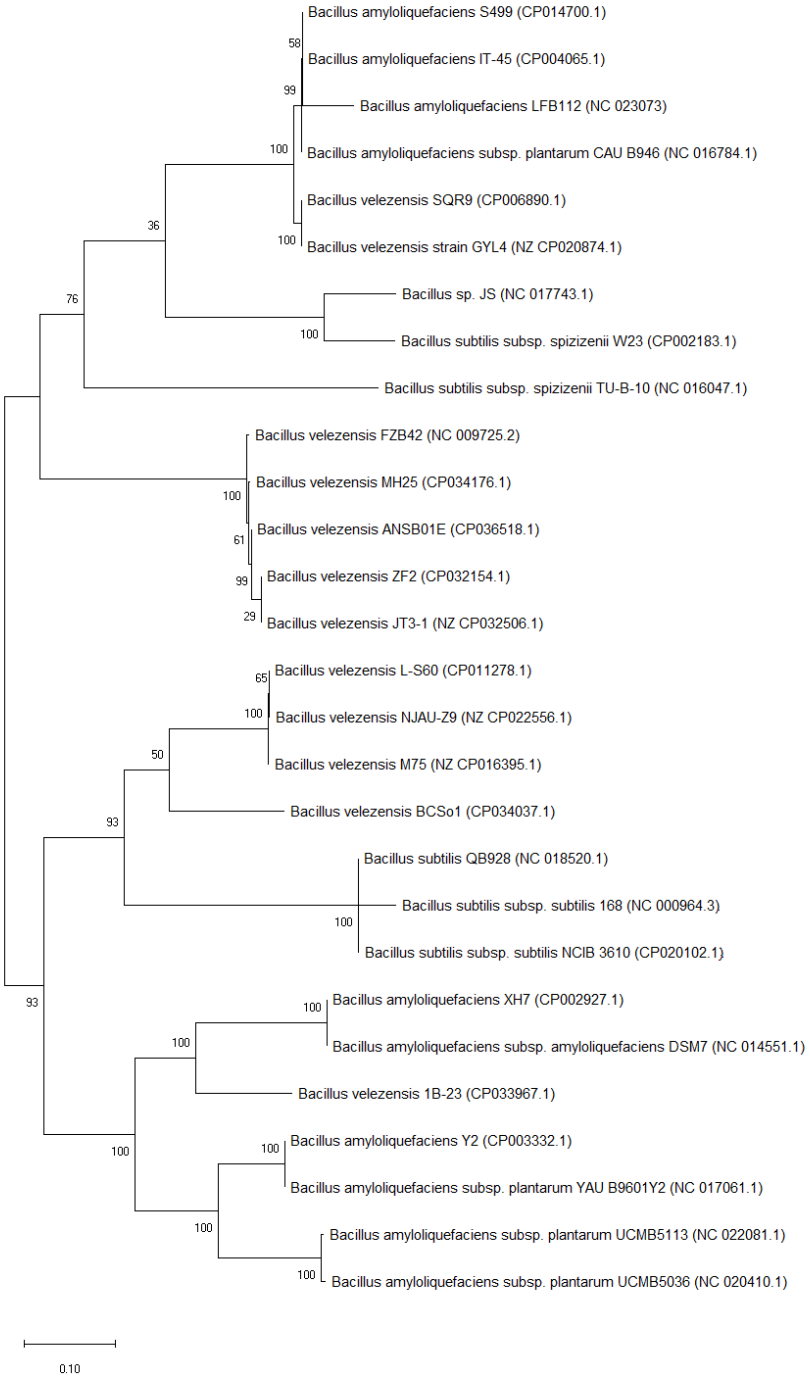


Figure 2.3. Phylogeny of the *comQXP* quorum sensing gene region for selected strains of *Bacillus subtilis sensu lato*. The evolutionary analysis was conducted in MEGA X using the Maximum Likelihood method and Kimura 2-parameter model. The tree represents the highest log likelihood (-33937.22). Phylogeny applied 28 nucleotide sequences with 3579 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

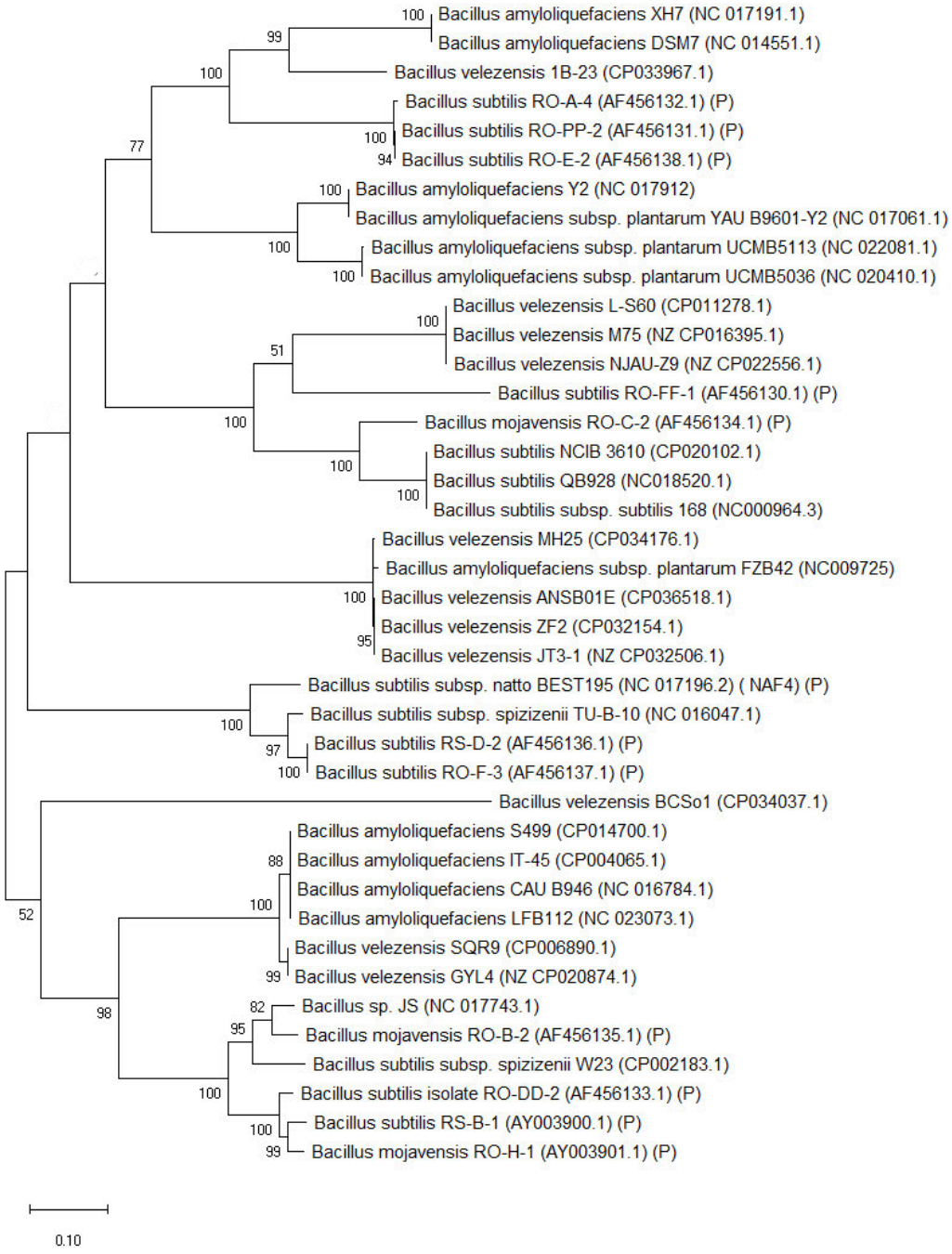


Figure 2.4. Phylogeny of the *comQ* quorum sensing gene region for selected strains of *Bacillus subtilis sensu lato* and known pherotypes (denoted (P)). The evolutionary analysis was conducted in MEGA X using the Maximum Likelihood method and Kimura 2-parameter model. The tree represents the highest log likelihood (-13928.63). Phylogeny applied 40 nucleotide sequences with 979 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

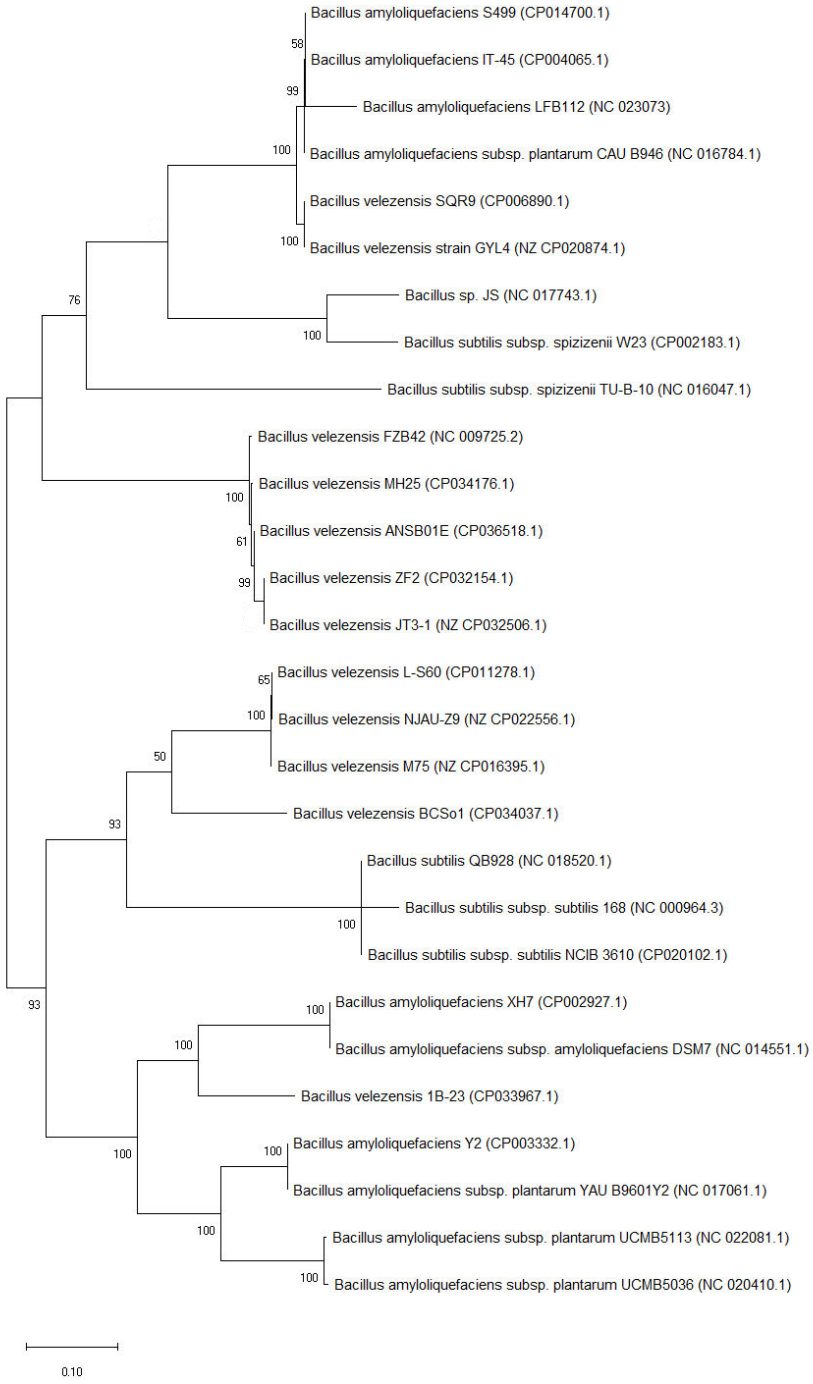


Figure 2.5. Phylogeny of simulated PCR amplicons of the *comQXP* quorum sensing gene region for in silico strains of *Bacillus subtilis sensu lato*. The evolutionary analysis was conducted in MEGA X using the Maximum Likelihood method and Kimura 2-parameter model. The tree represents the highest log likelihood (-29479.57). Phylogeny applied 32 nucleotide sequences with 3002 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

In silico PCR-RFLP of the simulated amplicons of the *comQXP* gene loci revealed several candidate restriction enzymes suitable for pherotype fingerprinting. These included *Cac*8I, *Bse*GI, *Bts*CI, *Fnu*4HI, *Ape*kI, *Hpy*CH4III, *Bst*KTI, *Dde*I, *Cvi*QI, *Hpy*166II, *Hpy*188I, *Mae*III and *Tfi*I (Data not shown). Four enzymes were selected as superior candidates: *Bts*CI, *Fnu*4HI, *Cac*8I, and *Hpy*166II. The projected fingerprints for these restriction enzymes have been shown in Figures 2.6. to 2.9., respectively. Simulated fingerprints illustrate enzyme suitability as per the criteria for selection. Briefly, these criteria were that fingerprints should be clear and easily readable, with the ability to distinguish *Bacillus* sp. JS from the wider *B. subtilis* group, and *B. amyloliquefaciens* subsp. *plantarum* (i.e. *B. velezensis*) from *B. amyloliquefaciens* subsp. *amyloliquefaciens*.

Comparing the strain clustering in the simulated *comQXP* phylogeny (Figure 2.6.) with the projected PCR-RFLP fingerprints (Figures 2.6.–2.9.), the restriction enzymes *Bts*CI, *Cac*8I, and *Hpy*166II simulate fingerprint variations which largely corroborate the phylogenetic clusters observed. There are instances of minor variations between the groupings noted in the enzymes themselves. For example, *B. amyloliquefaciens* strains DSM 7, LL3, TA208, and XH7 are clustered together with high confidence in the *comQXP* phylogeny and display identical fingerprints in the RFLP simulations of *Bts*CI, *Cac*8I and *Hpy*166II. However, *Fnu*4HI returns the same profile for DSM 7 and LL3, but these are distinct from the same profile of TA208 and XH7. Further examples of minor variation were observed in the *B. subtilis* strains: (1) *B. subtilis* subsp. *subtilis* RO-NN-1 displays a unique fingerprint for enzymes *Bts*CI, *Fnu*4HI, and *Hpy*166II, but the profile for *Cac*8I is nearly indistinguishable from that of *B. subtilis* strain XF-1 (This strain forms its own *com*QXP cluster in the phylogeny); and (2) *Bacillus* sp. JS forms a profile distinct from other strains for enzymes *Cac*8I, *Fnu*4HI, *Hpy*166II, but the *Bts*CI profile matches those of *B. subtilis* strains XF-1 and BAB-1. The *comQXP* phylogeny reveals *Bacillus* sp. JS to cluster with *B. subtilis* strains XF-1 and BAB-1.

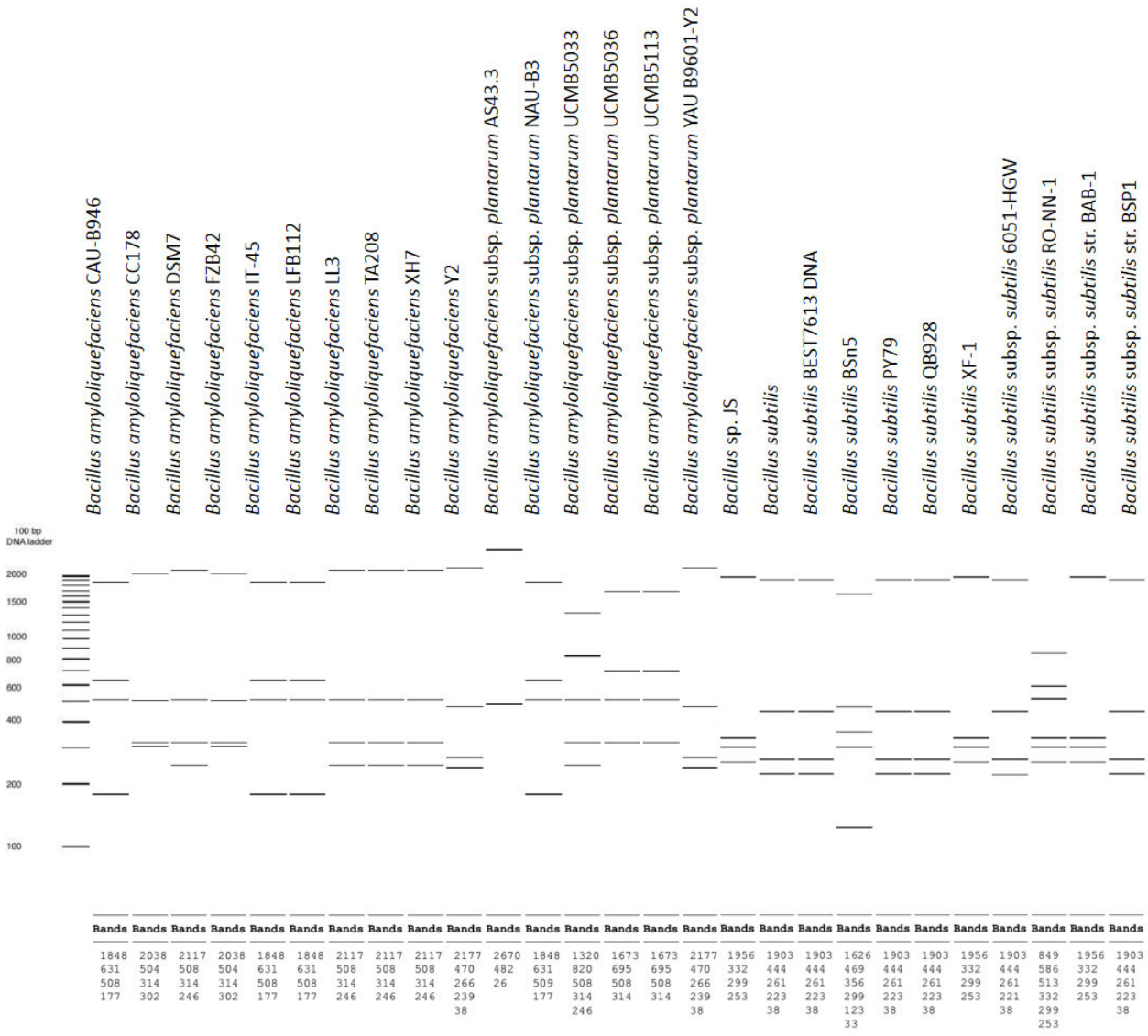


Figure 2.6. Simulated PCR-RFLP of the *comQXP* regions of *Bacillus amyloliquefaciens* and *B. subtilis* for restriction enzyme *Bts*CI. Fingerprints were generated from simulated amplicons of the *com*QXP quorum sensing region using the in silico PCR-RFLP tool (Bikandi *et al.*, 2004, San Millán *et al.*, 2013). The sizes (bp) of the bands in each profile are indicated underneath each lane.

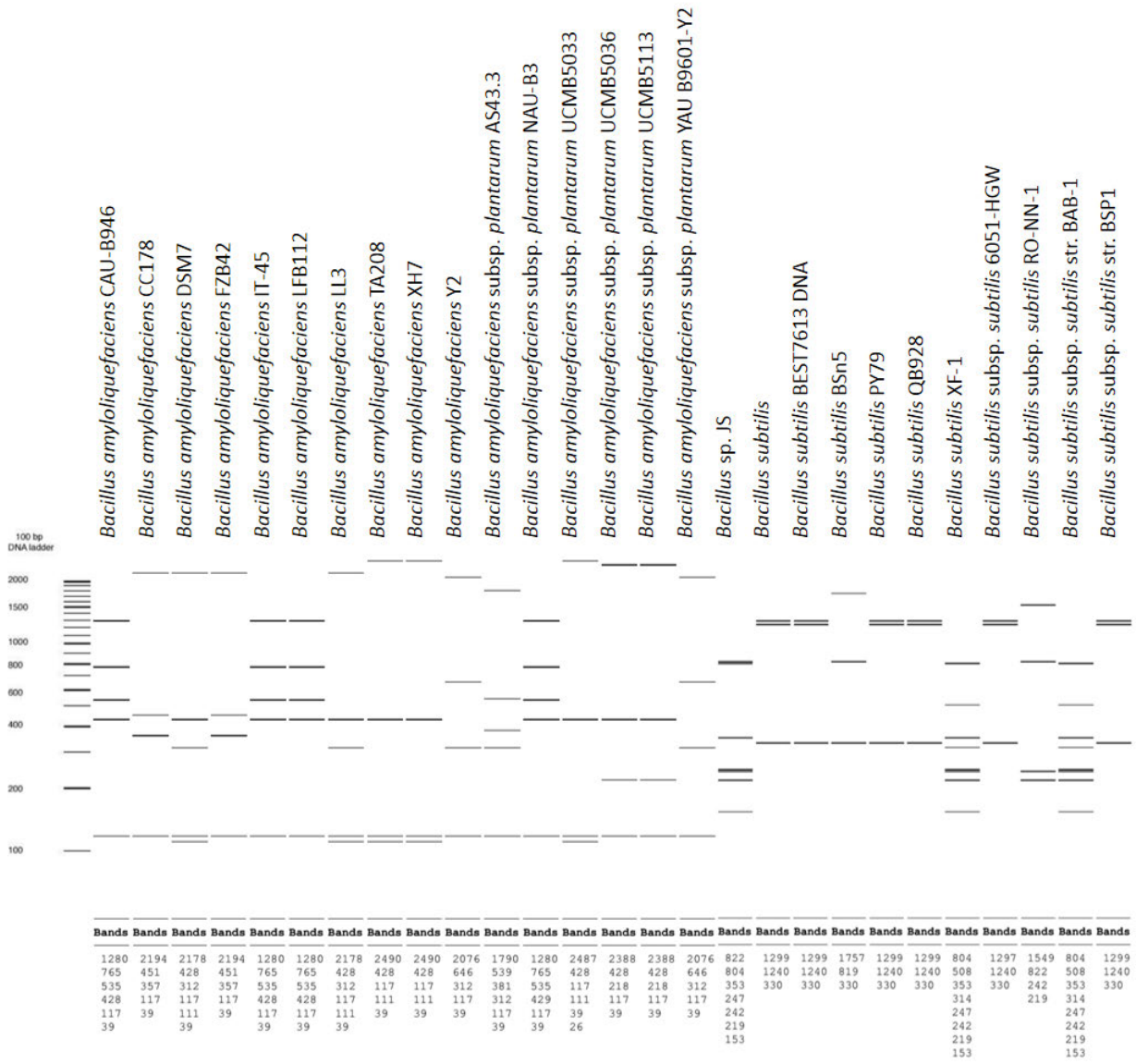


Figure 2.7. Simulated PCR-RFLP of the *comQXP* regions of *Bacillus amyloliquefaciens* and *B. subtilis* for restriction enzyme *Fnu*4HI. Fingerprints were generated from simulated amplicons of the *com*QXP quorum sensing region using the in silico PCR-RFLP tool (Bikandi *et al.*, 2004, San Millán *et al.*, 2013). The sizes (bp) of the bands in each profile are indicated underneath each lane.

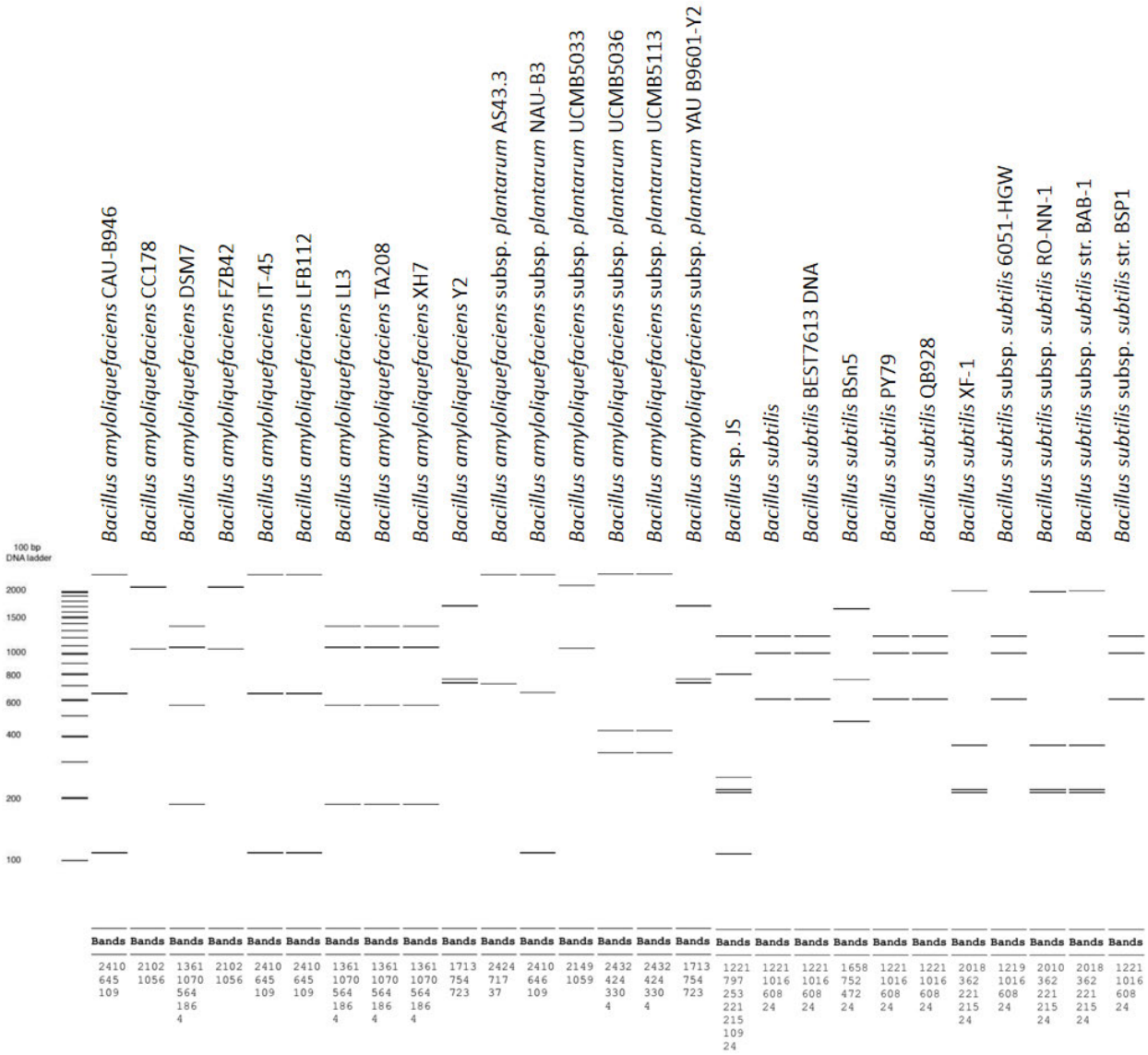


Figure 2.8. Simulated PCR-RFLP of the *comQXP* regions of *Bacillus amyloliquefaciens* and *B. subtilis* for restriction enzyme *Cac*8I. Fingerprints were generated from simulated amplicons of the *com*QXP quorum sensing region using the in silico PCR-RFLP tool (Bikandi *et al.*, 2004, San Millán *et al.*, 2013). The sizes (bp) of the bands in each profile are indicated underneath each lane.

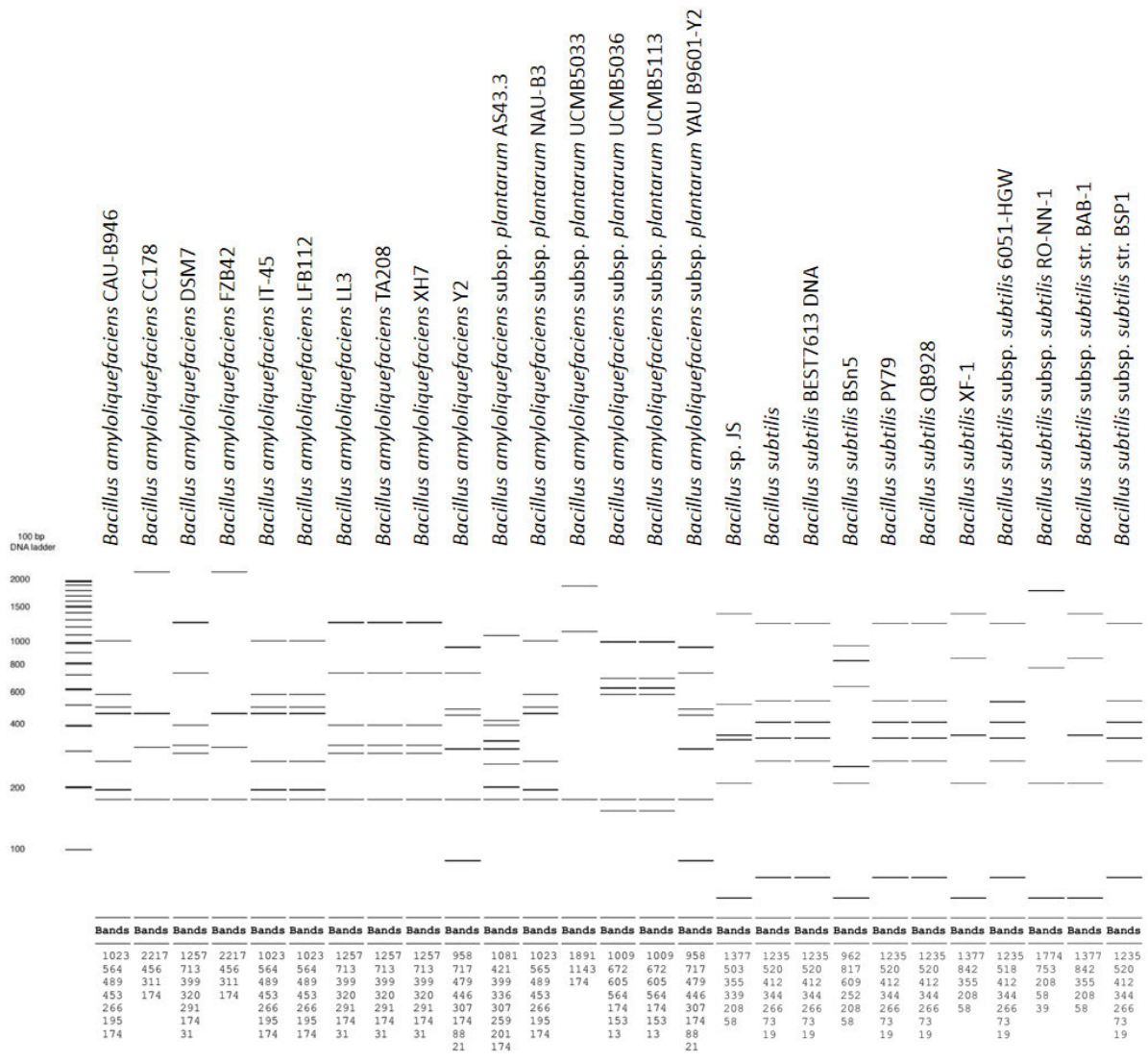


Figure 2.9. Simulated PCR-RFLP of the *comQXP* regions of *Bacillus amyloliquefaciens* and *B. subtilis* for restriction enzyme *Hpy*166II. Fingerprints were generated from simulated amplicons of the *com*QXP quorum sensing region using the in silico PCR-RFLP tool (Bikandi *et al.*, 2004, San Millán *et al.*, 2013). The sizes (bp) of the bands in each profile are indicated underneath each lane.

2.4. Discussion

This chapter details the use of in silico approaches to define subspecies variation and pherotypes amongst selected strains within the *Bacillus subtilis sensu lato* grouping, in particular strains of *B. subtilis* and *B. velezensis*. Both of the MLSA phylogenies resolved gene-based polymorphisms to subspecies level, while affording greater subgrouping resolution than was achievable by the conventional 16S rRNA gene sequence polymorphisms. This is not surprising, as a lack of subspecies level resolution is expected from phylogenies based on 16S rRNA gene sequence data, particularly in the *Bacillus* spp. (Mandić-Mulec *et al.*, 2015). The phylogenies of the individual housekeeping gene targets demonstrated broad similarity, with only minor sub-cluster variations as compared to either of the MLSA phylogenies. Some of these chosen gene sequences may be open to greater levels of polymorphism than others, which may explain this observed trend. Strain clustering across the individual gene sequence phylogenies was broadly consistent, with minor variations. Ultimately, the four genes selected (viz., 16S rRNA, *gyrA*, *dnaJ*, and *rpoB*) together generated a phylogeny that was the most consistent with the full-scale MLSA, and were sequences which had been applied previously—albeit individually—to studies of plant-associated *B. subtilis sensu lato* diversity (Rooney *et al.*, 2009, Stefanic and Mandić-Mulec, 2009, Borriss, 2011).

To the best of our knowledge, MLSA has not been applied to differentiate between the *B. amyloliquefaciens* species derivatives; in particular, the plant-associated *B. velezensis* of interest in the present study. The collective application of 16S rRNA, *gyrA*, *rpoB* and *dnaJ* generated a phylogeny that largely mirrored the clusters shown in the nine-gene MLSA and achieved very similar subspecies-level differentiation within these strains. The resolution of the proved sufficient to distinguish *B. velezensis* from B. *amyloliquefaciens*, and also generated a sub-cluster of *Bacillus* sp. JS from the major representatives *B. subtilis*, which is of interest as this plant-associated strain has demonstrated biocontrol potential (Song *et al.*, 2012). As such, the four-gene 16S rRNA-*gyrA*-*rpoB*-*dnaJ* MLSA supplied sufficient data for the purposes of this study.

Whole genome comparisons such as dDDH and ANI (Konstantinidis and Tiedje, 2005, Meier-Kolthoff *et al.*, 2013) are currently among the methods of choice to establish variation between strains and isolates. In broader studies with multiple isolates—such as those exploring microbial ecology—the PCR of individual targets demands time and resources, and whole genome sequencing of isolates is time-intensive and may not be financially viable for such high-throughput applications. MLSA affords a middle-ground solution, by applying several housekeeping gene sequences it is achievable with large subset of isolates. MLSA approaches have not yet been standardised for all species or applications, and gene selection can vary depending on application, though between five and seven genes are considered sufficiently robust for most MLSA applications (Ibarz Pavon and Maiden, 2009, Glaeser and Kämpfer, 2015). When selecting gene targets from whole genome sequence data—as in the present in silico study—one has access to a greater range of gene sequence targets.

The initial MLSA analysis undertaken in this in silico study targeted nine housekeeping genes. The application of MLSA to *Bacillus* spp. has not been standardised for any particular applications, and may comprise a range of gene targets. In previous studies applying MLSA to various *Bacillus* spp., there has been variability in the numbers and types of gene targets applied—between six and 25 (Bizzarri *et al.*, 2008, Kubo *et al.*, 2011, Liu *et al.*, 2013, Liu *et al.*, 2015, Carter *et al.*, 2018, Le *et al*., 2019). The number of gene targets selected will impact the resolving capability of an MLSA, and it becomes a matter of compromise between the level of discrimination required and the cost of gene sequencing (Glaeser and Kämpfer, 2015). As such, this study started with a larger dataset and used that information to build a smaller phylogeny while preserving a comparable level of subspecies resolution. In cases where high-throughput is valued, the MLSA parameters may be adjusted to best meet the specific demands of the study undertaken, such as lowering the number of housekeeping gene targets, as was undertaken in the present study.

Pherotyping in *Bacillus* spp. may be achieved either by direct study of the ComX pheromones themselves, or analysis of the *comQXP* quorum sensing loci or portions thereof. In this study pherotyping of *B. subtilis sensu lato* made use of quorum-sensing loci gene sequence data from GenBank and from simulated PCR amplicons. The *comQXP* gene sequence phylogeny

generated for this study had taken advantage of the communication loci regions in direct contact with the hypervariable ComX (viz., *comQ*, *comX*, and the portion of *comP* involved in cognate ComX binding) (Tran *et al.*, 2000, Tortosa *et al.*, 2001, Ansaldi *et al.*, 2002). Given the size of this particular gene region (>2 kb), many studies only sequence the *comQ* portion of the amplicon and consider this sufficient to indicate pherotype variation (Ansaldi *et al.*, 2002, Stefanic and Mandić-Mulec, 2009, Stefanic *et al.*, 2012, Ozlislo *et al.*, 2015). In this study, the *comQ* phylogeny showed sufficient similarity to the *comQXP* phylogeny to support its use as a sequence-based pherotyping technique.

In both quorum-sensing sequence phylogenies, inter- and intra- species variation in the DNA sequences of these quorum sensing loci was observed. The clusters observed in both *comQ* and *comQXP* phylogenies indicate degrees of inter-species similarities in these gene sequences, which may be linked to horizontal gene transfer between closely-related *B. subtilis sensu lato* members, as indicated by the lower relative G+C content of this region (Tortosa and Dubnau, 1999, Mandić-Mulec *et al.*, 2003, Eldar, 2011). Furthermore, marked sequence variation was apparent between the selected strains and established pherotypes, suggesting significant quorum-sensing divergence, in particular amongst *B. velezensis* strains. ComX is known to serve as an intra-population means of communication and therefore is under constant evolutionary pressure to protect so-called “public goods” (i.e. the beneficial products of social behaviours) through the continual modification of discrete communication pheromones (Stefanic and Mandić-Mulec, 2009). As such, it would not be unexpected for plant-adapted *B. velezensis* to demonstrate unique ComX variants.

Several putative pherotype clusters were observed within the *B. velezensis* representatives in this study. With regards to the animal-associated strains (viz., *B. velezensis* ANSB01E, JT3-1, BCSo1), it is possible that the bacterium was taken in by the animal via interactions with the environment, such as dietary plant matter. The sequence clustering of the majority of the most widely-used of the *B. subtilis* laboratory strains (viz., *B. subtilis* 168, QB928, NCIB3610) demonstrated a limited divergence in the quorum-sensing loci of these strains, perhaps owing to a lack of selection pressure for discrete communication by the absence of niche competitors and gene transfer events occurring in the laboratory environment. Comparison of the clustering

within the phylogenetic trees generated for the gene sequence data of the selected *B. subtilis sensu lato* strains in the four-gene MLSA and *comQXP* suggests that there exists pherotype divergence amongst the MLSA clusters, indicating that closely-related isolates may harbour distinct ComX variants.

Simulated *comQXP* PCR-RFLP demonstrated that this approach could be suitable for high-throughput pherotyping. This study endeavoured to select restriction enzyme candidates which were able to distinguish some of the very-closely related species in the *B. subtilis sensu lato* grouping; namely *B. velezensis* (syn. *B. amyloliquefaciens* subsp. *plantarum*) from the *B. amyloliquefaciens* subspecies, and plant-adapted *Bacillus* sp. JS from the *B. subtilis sensu stricto*. The counterpart simulated PCR *comQXP* phylogeny broadly reflected the groupings observed in the PCR-RFLPs. A suite of restriction enzymes should always be applied to PCR-RFLP in order to ensure that the diversity observed is as representative as possible (Li *et al.*, 2009), which was found to be the case in the present study. Four enzyme candidates were presented (viz., *Cac*8I, *Bts*CI, *Hpy*166II, and *Fnu*4HI). Minor PCR-RFLP fingerprint profile variations were apparent with some enzymes, which may be explained by the presence of non-specific cut sites (N-nucleobases) of certain restriction enzymes. These restriction enzymes together proved able to reflect fingerprint polymorphisms which mirrored the clustering observed in the *comQXP* tree, with acceptable minor deviations.

2.4.1. Conclusions

The sequence-based experiments presented in this study indicate that MLSA has potential a useful approach towards defining subspecies-level amongst a subset of *B. subtilis sensu lato* strains, in particular those strains designated as plant-associated *B. velezensis*. Ecological variation amongst members of the *B. subtilis sensu lato* has been reported, and putative bacterial ecotypes are frequently phylogenetically differentiated on the basis of sequence clusters of housekeeping genes (Koeppel *et al.*, 2008, Dogsa *et al.*, 2014) such as those achievable using the MLSA approach (Cohan, 2002, Maiden, 2006). In silico housekeeping gene sub-clustering and pherotype data suggested that pherotype variability may be relevant to plant-adapted ecotype, further assessments applying a larger dataset and Ecotype Simulation

analyses will provide more rigorous supporting evidence and better resolution of ecotype-level polymorphisms.

In this study, subspecies-level strain clustering observed in the MLSA did not consistently mirror clustering in the quorum-sensing loci phylogeny, which indicated that there may exist levels of ComX diversity even amongst closely-related strains. To what degree ComX divergence plays a role in subspecies and ecotype variation remains to be further expanded upon, but *comQXP* and *comQ* DNA sequence analyses did suggest a strong possibility of the existence of novel pherotypes in *B. velezensis* and *B. amyloliquefaciens*. Investigations into the pherotypes of environmental *Bacillus subtilis sensu lato* are few (Ansaldi *et al.*, 2002, Oslizlo *et al.*, 2015, Esmaeilishirazifard *et al.*, 2017) and the degree of ComX communication divergence throughout the genus is relatively unexplored, even outside of the plant-associated focus of the present study. As an indirect sequence-based pherotyping approach, the in silico *comQXP* PCR-RFLP showed great promise as a means of *B. subtilis sensu lato* spp. pherotyping, but requires experimental validation utilising reference strains to confirm projected fingerprint similarities.

2.5. References

Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L. & Olmedo, G. (2010) Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* **11**: 332, 1–17. 10.1186/1471-2164-11-332

Ansaldi, M. & Dubnau, D. (2004) Diversifying selection at the *Bacillus* quorum-sensing locus and determinants of modification specificity during synthesis of the ComX pheromone. *Journal of Bacteriology* **186**: 1, 15–21. 10.1128/jb.186.1.15-21.2004

Ansaldi, M., Marolt, D., Stebe, T., Mandić-Mulec, I. & Dubnau, D. (2002) Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Molecular Microbiology* **44**: 6, 1561–1573. https://doi.org/10.1046/j.1365-2958.2002.02977.x

Bikandi, J., San Millán, R., Rementeria, A. & Garaizar, J. (2004) *In silico* analysis of complete bacterial genomes: PCR, AFLP–PCR and endonuclease restriction. *Bioinformatics* **20**: 5, 798-799. 10.1093/bioinformatics/btg491

Bizzarri, M. F., Prabhakar, A. & Bishop, A. H. (2008) Multiple-Locus Sequence Typing Analysis of Bacillus thuringiensis Recovered from the Phylloplane of Clover (Trifolium hybridum) in Vegetative Form. *Microbial Ecology* **55**: 4, 619-625. 10.1007/s00248-007-9305-3

Borriss, R. (2011) Use of plant-associated *Bacillus* strains as biofertilizers and biocontrol agents in agriculture. *Bacteria in Agrobiology: Plant Growth Responses* (Maheshwari, D. K., ed.) pp. 41–76. Springer-Verlag, Berlin, Germany. 10.1007/978-3-642-20332-9_3

Carter, L., Chase, H. R., Gieseker, C. M., Hasbrouck, N. R., Stine, C. B., Khan, A., Ewing-Peeples, L. J., Tall, B. D. & Gopinath, G. R. (2018) Analysis of enterotoxigenic Bacillus cereus strains from dried foods using whole genome sequencing, multi-locus sequence analysis and toxin gene prevalence and distribution using endpoint PCR analysis. *International Journal of Food Microbiology* **284**: 31-39. https://doi.org/10.1016/j.ijfoodmicro.2018.06.016

Chen, X. H., Koumoutsi, A., Scholz, R., Schneider, K., Vater, J., Süssmuth, R., Piel, J. & Borriss, R. (2009) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *Journal of Biotechnology* **140**: 27–37. 10.1016/j.jbiotec.2008.10.011

Cohan, F. M. (2002) What are bacterial species? *Annual Reviews in Microbiology* **56**: 457–487. 10.1146/annurev.micro.56.012302.160634

Cohan, F. M. (2005) Periodic selection and ecological diversity in bacteria. *Selective Sweep* (Nurminsky, D., ed.) pp. 78–93. Springer US, Boston, MA. 10.1007/0-387-27651-3_7

Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koeppel, A. F., Burger, A., Cole, S. G., Perry, E. B., Krizanc, D., Field, N. C., Slaton, M. & Cohan, F. M. (2010) Ecology of speciation in the genus *Bacillus*. *Applied and Environmental Microbiology* **76**: 5, 1349–1358. 10.1128/aem.01988-09

Dedonder, R. A., Lepesant, J. A., Lepesant-Kejzlarová, J., Billault, A., Steinmetz, M. & Kunst, F. (1977) Construction of a kit of reference strains for rapid genetic mapping in *Bacillus subtilis* 168. *Applied and Environmental Microbiology* **33**: 4, 989-993. 10.1128/AEM.33.4.989-993.1977

Dogsa, I., Oslizlo, A., Stefanic, P. & Mandić-Mulec, I. (2014) Social interactions and biofilm formation in *Bacillus subtilis*. *Food Technology and Biotechnology* **52**: 2, 149–157. 10.1128/JB.01290-08

Dunlap, C. A., Kim, S.-J., Kwon, S.-W. & Rooney, A. P. (2015) Phylogenomic analysis shows that *Bacillus amyloliquefaciens* subsp. *plantarum* is a later heterotypic synonym of *Bacillus methylotrophicus*. *International Journal of Systematic and Evolutionary Microbiology* **65**: 7, 2104–2109. doi:10.1099/ijs.0.000226

Dunlap, C. A., Kim, S.-J., Kwon, S.-W. & Rooney, A. P. (2016) *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and *'Bacillus oryzicola'* are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. *International Journal of Systematic and Evolutionary Microbiology* **66**: 3, 1212–1217. doi:10.1099/ijsem.0.000858

Earl, A. M., Losick, R. & Kolter, R. (2008) Ecology and genomics of *Bacillus subtilis*. *Trends in Microbiology* **16**: 6, 269-275. 10.1016/j.tim.2008.03.004

Edgar, R. C. (2004) MUSCLE: Multiple alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 5, 1792-1797. 10.1093/nar/gkh340

Eldar, A. (2011) Social conflict drives the evolutionary divergence of quorum sensing. *Proceedings of the National Academy of Sciences Of the United States of America* **108**: 33, 13635–13640. 10.1073/pnas.1102923108

Esmaeilishirazifard, E., De Vizio, D., Moschos, S. A. & Keshavarz, T. (2017) Genomic and molecular characterization of a novel quorum sensing molecule in *Bacillus licheniformis*. *AMB Express* **7**: 78, 1–13. 10.1186/s13568-017-0381-6

Fajardo-Cavazos, P., Maughan, H. & Nicholson, W. L. (2014) Evolution in the *Bacillaceae*. *Microbiology Spectrum* **2**: 5, 1–32. 10.1128/microbiolspec.TBS-0020-2014

Fan, B., Blom, J., Klenk, H.-P. & Borriss, R. (2017) *Bacillus amyloliquefaciens*, *Bacillus velezensis* and *Bacillus siamensis* form an 'operational group *B. amyloliquefaciens*' within the *B. subtilis* species complex. *Frontiers in Microbiology* **8**: 22, 1–15. 10.3389/fmicb.2017.00022

Fan, B., Wang, C., Ding, X., Zhu, B., Song, X. & Borriss, R. (2019) AmyloWiki: An integrated database for *Bacillus velezensis* FZB42, the model strain for plant growth-promoting *Bacilli*. *Database: The Journal of Biological Databases and Curation* **2019**: baz071. 10.1093/database/baz071

Fritze, D. (2004) Taxonomy of the genus *Bacillus* and related genera: The aerobic endospore-forming bacteria. *Phytopathology* **94**: 11, 1245–1248. https://doi.org/10.1094/PHYTO.2004.94.11.1245

Glaeser, S. P. & Kämpfer, P. (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology* **38**: 4, 237–245. http://dx.doi.org/10.1016/j.syapm.2015.03.007

Hall, T. A. (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98. 10.14601/Phytopathol_Mediterr-14998u1.29

Ibarz Pavon, A. B. & Maiden, M. C. (2009) Multilocus sequence typing. *Molecular Epidemiology of Microorganisms Methods in Molecular Biology (Methods and Protocols)* Vol. 551 (Caugant, D., ed.) pp. 129–140. Humana Press, Totowa, NJ, USA. 10.1007/978-1-60327-999-4_11

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 2, 111-120. 10.1007/bf01731581

Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., Rooney, A. P., Brambilla, E., Connor, N., Ratcliff, R. M., Nevo, E. & Cohan, F. M. (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 7, 2504–2509. 10.1073/pnas.0712205105

Konstantinidis, K. T. & Tiedje, J. M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 7, 2567–2572. 10.1073/pnas.0409727102

Kopac, S., Wang, Z., Wiedenbeck, J., Sherry, J., Wu, M. & Cohan, F. M. (2014) Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. *Applied and Environmental Microbiology* **80**: 16, 4842–4853. 10.1128/aem.00576-14

Kubo, Y., Rooney, A. P., Tsukakoshi, Y., Nakagawa, R., Hasegawa, H. & Kimura, K. (2011) Phylogenetic analysis of *Bacillus subtilis* strains applicable to natto (fermented soybean) production. *Applied and Environmental Microbiology* **77**: 18, 6463-6469. 10.1128/aem.00448-11

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 6, 1547-1549. 10.1093/molbev/msy096

Le, X. T., Pham, D. T., Pham, T. A., Tran, T. T., Khuat, T. H., Le, H. Q. & Vu, U. N. (2019) Exploration of genetic diversity of Bacillus spp. from industrial shrimp ponds in Vietnam by multi-locus sequence typing. *Fisheries and Aquatic Sciences* **22**: 1, 17. 10.1186/s41240-019-0132-5

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G. & Shao, Z. (2013) Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. *PLoS One* **8**: 11, e80097. 10.1371/journal.pone.0080097

Liu, Y., Lai, Q., Göker, M., Meier-Kolthoff, J. P., Wang, M., Sun, Y., Wang, L. & Shao, Z. (2015) Genomic insights into the taxonomic status of the Bacillus cereus group. *Scientific Reports* **5**: 1, 14082. 10.1038/srep14082

Maiden, M. C. (2006) Multilocus sequence typing of bacteria. *Annual Review of Microbiology* **60**: 561–588. 10.1146/annurev.micro.59.030804.121325

Mandić-Mulec, I., Stefanic, P. & van Elsas, J. D. (2015) Ecology of *Bacillaceae*. *Microbiology Spectrum* **3**: 1, 1-23. 10.1128/microbiolspec.TBS-0017-2013

Mandić-Mulec, I., Kraigher, B., Cepon, U. & Mahne, I. (2003) Variability of the quorum sensing system in natural isolates of *Bacillus* sp. *Food Technology and Biotechnology* **47**: 1, 23–28.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**: 60. 10.1186/1471-2105-14-60

Nishito, Y., Osana, Y., Hachiya, T., Popendorf, K., Toyoda, A., Fujiyama, A., Itaya, M. & Sakakibara, Y. (2010) Whole genome assembly of a natto production strain Bacillus subtilis natto from very short read data. *BMC Genomics* **11**: 243-243. 10.1186/1471-2164-11-243

Oslizlo, A., Stefanic, P., Vatovec, S., Beigot Glaser, S., Rupnik, M. & Mandić-Mulec, I. (2015) Exploring ComQXPA quorum-sensing diversity and biocontrol potential of *Bacillus* spp. isolates from tomato rhizoplane. *Microbial Biotechnology* **8**: 3, 527–540. 10.1111/1751-7915.12258

Reva, O. N., Dixelius, C., Meijer, J. & Priest, F. G. (2004) Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *FEMS Microbiology Ecology* **48**: 249–259. 10.1016/j.femsec.2004.02.003

Rooney, A. P., Price, N. P. J., Ehrhardt, C., Swezey, J. L. & Bannan, J. D. (2009) Phylogeny and molecular taxonomy of the *Bacillus subtilis* species complex and description of *Bacillus subtilis* subsp. *inaquosorum* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology* **59**: 2429–2436. 10.1099/ijs.0.009126-0

San Millán, R. M., Martínez-Ballesteros, I., Rementeria, A., Garaizar, J. & Bikandi, J. (2013) Online exercise for the design and simulation of PCR and PCR-RFLP experiments. *BMC Research Notes* **6**: 1, 513. 10.1186/1756-0500-6-513

Song, J. Y., Kim, H. A., Kim, J.-S., Kim, S.-Y., Jeong, H., Kang, S. G., Kim, B. K., Kwon, S.-K., Lee, C. H., Yu, D. S., Kim, B. S., Kim, S.-H., Kwon, S. Y. & Kim, J. F. (2012) Genome sequence of the plant growth-promoting rhizobacterium *Bacillus* sp. strain JS. *Journal of Bacteriology* **194**: 14, 3760-3761. 10.1128/jb.00676-12

Stefanic, P. & Mandić-Mulec, I. (2009) Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. *Journal of Bacteriology* **191**: 6, 1756–1764. 10.1128/jb.01290-08

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F. M. & Mandić-Mulec, I. (2012) The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. *Environmental Microbiology* **14**: 6, 1378–1389. 10.1111/j.1462-2920.2012.02717.x

Tortosa, P. & Dubnau, D. (1999) Competence for transformation: A matter of taste. *Current Opinion in Microbiology* **2**: 6, 588–592. https://doi.org/10.1016/S1369-5274(99)00026-0

Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandić-Mulec, I. & Dubnau, D. (2001) Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. *Journal of Bacteriology* **183**: 2, 451–460. 10.1128/JB.183.2.451-460.2001

Tran, L. S., Nagai, T. & Itoh, Y. (2000) Divergent structure of the ComQXPA quorum-sensing components: Molecular basis of strain-specific communication mechanism in *Bacillus subtilis*. *Molecular Microbiology* **37**: 5, 1159–1171. 10.1046/j.1365-2958.2000.02069.x

Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kuhl, M. & Grossman, A. (2007) Genomics, environmental genomics and the issue of microbial species. *Heredity* **100**: 2, 207–219. 10.1038/sj.hdy.6801011

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 1, 134–145. 10.1186/1471-2105-13-134

Yu, C.-S., Yim, K.-Y., Tsui, S. K.-W. & Chan, T.-F. (2012) Complete genome sequence of *Bacillus subtilis* strain QB928, a strain widely used in *B. subtilis* genetic studies. *Journal of Bacteriology* **194**: 22, 6308–6309. 10.1128/jb.01533-12

CHAPTER THREE

ESTABLISHING ECOTYPE VARIATION AMONGST PLANT-ASSOCIATED *BACILLUS SUBTILIS SENSU LATO* ISOLATES

3.1. Introduction

Bacteria are known to possess adaptations that allow them to expand into and colonise specific ecological niches (Lan and Reeves, 2000). These adaptations arise from environmental selection pressures, and result in functional divergence within the populations of species (Cohan, 2005, Ward *et al.*, 2007, Koeppel *et al.*, 2008). Such ecologically adapted species populations are termed ecotypes (Cohan, 2005). Accounting for ecotype variation within conventional bacterial taxonomy has been widely debated in recent years (Ward *et al.*, 2007, Cohan and Koeppel, 2008, Fraser *et al.*, 2009, Wiedenbeck and Cohan, 2011, Lassalle *et al.*, 2015). The ecological species concept aims to reconcile a prokaryote's genetic information with its environmental niche, and examines the associated functional genes determining the organism's success in that community (Ward, 2002). Functional genes are often more highly varied in DNA sequence between species than the benchmark 16S rRNA gene sequence may infer (Ward, 2002, Cohan, 2006). Ecotypes could be considered as fundamental units of prokaryotic diversity, and the ecological species concept may be crucial to understanding microbial community composition, structure, and function (Koeppel *et al*., 2008). The ability to accurately describe an ecotype further provides a standard for grouping of diversity that is useful to many disciplines where bacterial ecology and evolution intersect (Ward, 2006, Koeppel *et al.*, 2008).

The ubiquitous genus *Bacillus* has featured prominently in ecological species debates, as this genus comprises phenotypically heterogeneous bacteria occupying a range of environmental niches, with an equally varied and complicated taxonomy (Fritze, 2004, Schleifer, 2009, Fajardo-Cavazos *et al.*, 2014, Mandić-Mulec *et al.*, 2015). *Bacillus* spp. housekeeping gene heterogeneity, coupled with functional diversity within members, has led to the description of *Bacillus* species complexes (Fritze, 2004). These groups of related taxa—*sensu lato*

groupings—account for the high degree of variance in genomic data between closely-related species, and this also reflects their divergence in ecological terms. Ecologically distinct groupings (i.e. ecotypes) amongst *B. subtilis, B. licheniformis,* and *B. thuringiensis* species have been described in several studies (Reva *et al.*, 2004, Koeppel *et al.*, 2008, Alcaraz *et al.*, 2010, Connor *et al.*, 2010, Stefanic *et al.*, 2012, Swiecicka *et al.*, 2013, Kopac *et al.*, 2014). Ecological adaptation amongst plant-associated *B. subtilis sensu lato* members has been previously reported, and includes plant-health-promoting strains with commercialization potential (Cawoy *et al.*, 2011, Zhang *et al.*, 2016, Dunlap, 2019, Reva *et al.*, 2019). *Bacillus velezensis*, in particular, has attracted significant research attention for its colonisation abilities and multiple beneficial activities in plant health (Chen *et al.*, 2009a, Chen *et al.*, 2009b, Fan *et al.*, 2012). Further, *B. velezensis* serves as a valuable example of the influence of ecological adaptation on bacterial taxonomy. Despite the plant-beneficial activities associated with this species, to the best of our knowledge, ecotype variation amongst plant-associated *B. velezensis* populations has not been studied. Identifying ecotypes within plant-growth-promoting bacteria, such as those in the *B. subtilis sensu lato*, is useful to further our appreciation of characteristics necessary for niche adaptation which may be targeted in environmental isolate selection procedures, as well as understanding just how these could impact niche-specialisation and habitat persistence. Ecotype knowledge may also provide indications of strain adaptation to prevailing local conditions and plant types, for targeted biocontrol and biofertilizer applications.

Previous studies have reconciled functional and ecological variance amongst closely related bacterial species using approaches such as Multilocus Sequence Analysis (MLSA) (Ibarz Pavon and Maiden, 2009); PCR-based fingerprinting approaches (e.g. Rep-PCR, RAPD-PCR) (Istock *et al.*, 2001); and sequence-based determination of ecotypes using the Ecotype Simulation (ES) algorithm derived by Koeppel *et al.* (2008). In the present study, these methods were applied to ascertain the level of ecotype diversity amongst *Bacillus subtilis sensu lato* isolates derived from crop plant rhizosphere and phyllosphere samples.

3.2. Materials and Methods

3.2.1 Bacterial strains

The plant-associated bacterial isolates used in this study comprised a total of 24 strains of *B. subtilis* and *B. velezensis* (Table 3.1) identified previously on the basis of gyrase subunit alpha (*gyrA*) gene sequence data (Tredgold, 2015, Hunter, 2016) (Table B1). Details of isolates origins are presented in previous works (Tredgold, 2015, Hunter, 2016). As detailed in Table 3.1., *Bacillus* spp. isolates were isolated as endospores from their respective environments and crop species. In some cases, plants species are represented in multiple instances, dependent on their availability across the different sampling locations. The isolates selected for inclusion in this study had shown some degree of antifungal activity and plant-growth promotion characteristics which were determined previous studies using both in vitro approaches and pot trials (Tredgold, 2015, Hunter, 2016).

Isolates were confirmed as *B. velezensis* strains using endpoint PCR targeting a sugar kinase unique to *B. velezensis* (GenBank RBAM_RS08965) (Dunlap, 2019) (Figure B1). Type strains *B. subtilis* subsp. *subtilis* DSM 10^T (=Marburg), *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T (=Fukumoto strain F), and *B. velezensis* DSM 23117^T (=FZB42) (Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany) were also included for comparative purposes. Counterpart housekeeping gene sequences from pertinent *B. subtilis sensu lato* reference genomes were included for phylogenetic analyses. These sequences were obtained from the NCBI GenBank database, and are listed in Table 3.2. The outgroup sequence for phylogenetic analyses was *B. cereus* ATCC 14579 (NC_004722.1).

Table 3.1. *Bacillus subtilis sensu lato* isolates evaluated in this study. Isolates originated from phyllosphere and rhizosphere samples from various crop plants in the KwaZulu-Natal province, South Africa).

Isolate	Species	NCBI GenBank Accession Number				Plant Species and Locality	Reference
		16S rRNA	*gyr*A	*rpo*B	*dna*J		
bna81	*B. velezensis*	MW237719	MW254234	MW254277	MW254253	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
ccc103	*B. velezensis*	MW237720	MW254235	MW254278	MW254254	Cucumber leaf, KwaZulu-Natal	Tredgold (2015)
bnd137	*B. velezensis*	MW237723	MW254238	MW254281	MW254257	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
cce142	*B. velezensis*	MW237724	MW254239	MW254282	MW254258	Cucumber leaf, KwaZulu-Natal	Tredgold (2015)
cce146	*B. velezensis*	MW237725	MW254240	MW254283	MW254259	Cucumber leaf, KwaZulu-Natal	Tredgold (2015)
bnd160	*B. velezensis*	MW237727	MW254242	MW254285	MW254261	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng199	*B. velezensis*	MW237728	MW254243	MW254286	MW254262	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng230	*B. velezensis*	MW237731	MW254246	MW254289	MW254265	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
pkl242	*B. velezensis*	MW237732	MW254247	MW254290	MW254266	Pumpkin leaf, KwaZulu-Natal	Tredgold (2015)
sqo275	*B. velezensis*	MW237734	MW254249	MW254292	MW254268	Squash leaf, KwaZulu-Natal	Tredgold (2015)
sqo277	*B. velezensis*	MW237735	MW254250	MW254293	MW254269	Squash leaf, KwaZulu-Natal	Tredgold (2015)
bnn282	*B. velezensis*	MW237737	MW254252	MW254295	MW254271	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
R16	*B. velezensis*	KT920463.1	KT960039.1	MW254296	MW254272	Lettuce root, KwaZulu-Natal (Medium: soil)	Hunter (2016)
CP-R23	*B. velezensis*	KT920456.1	KT960037.1	MW254297	MW254273	Capsicum root, KwaZulu-Natal (Medium: Composted pine bark)	Hunter (2016)
CL-R51	*B. velezensis*	KT920452.1	KT960038.1	MW254298	MW254274	Lettuce root, KwaZulu-Natal (Medium: Composted pine bark)	Hunter (2016)
bnd134	*B. subtilis*	MW237721	MW254236	MW254279	MW254255	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bnd136	*B. subtilis*	MW237722	MW254237	MW254280	MW254256	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bnd156	*B. subtilis*	MW237726	MW254241	MW254284	MW254260	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng215	*B. subtilis*	MW237729	MW254244	MW254287	MW254263	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng221	*B. subtilis*	MW237730	MW254245	MW254288	MW254264	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
sqo271	*B. subtilis*	MW237733	MW254248	MW254291	MW254267	Squash leaf, KwaZulu-Natal	Tredgold (2015)
sqo279	*B. subtilis*	MW237736	MW254251	MW254294	MW254270	Squash leaf, KwaZulu-Natal	Tredgold (2015)
CT-R67	*B. subtilis*	KT920460.1	KT960045.1	MW254299	MW254275	Tomato root, KwaZulu-Natal (Medium: Composted pine bark)	Hunter (2016)
B81	*B. subtilis*	KT920447.1	KT960043.1	MW254300	MW254276	Pumpkin roots, KwaZulu-Natal (Medium: soil)	Hunter (2016)

Table 3.2. Representative genomes of *Bacillus subtilis sensu lato* selected for this study sourced from the NCBI GenBank database.

Species and Strain	GenBank Accession No.	Origin / Source
B. velezensis FZB42	NC_009725.2	Beet field soil
B. velezensis SQR9	CP006890.1	Cucumber rhizosphere soil
B. velezensis L-S60	CP011278.1	Soil
B. velezensis GYL4	NZ_CP020874.1	Pepper plant
B. velezensis NJAU-Z9	NZ_CP022556.1	Field soil
B. velezensis M75	NZ_CP016395.1	Mushroom cultivation substrate (Cotton waste)
B. velezensis MH25	CP034176.1	Rhizosphere soil
B. velezensis BCSo1	CP034037.1	*Danio rerio*
B. velezensis 1B-23	CP033967.1	Potato rhizosphere
B. velezensis ANSB01E	CP036518.1	Chicken intestine
B. velezensis ZF2	CP032154.1	Cucumber plants
B. velezensis JT3-1	NZ_CP032506.1	Yak faeces
B. amyloliquefaciens S499	CP014700.1	Plant soil
B. amyloliquefaciens subsp. *amyloliquefaciens* DSM 7	NC_014551.1	DSMZ[¥] type strain
B. amyloliquefaciens subsp. *plantarum* UCMB5036	NC_020410.1	Inner tissues cotton plant (*Gossypium barbadense*)
Bacillus sp. JS	NC_017743.1	*Miscanthus* spp. soil
B. subtilis subsp. *spizizenii* TU-B-10	NC_016047.1	Tunisian desert
B. subtilis subsp. *spizizenii* W23	CP002183.1	BGSC* laboratory reference strain
B. subtilis subsp. *inaquosorum* KCTC 13429	CP029465.1	Soil (Death Valley, California, U. S. A.)
B. subtilis subsp. *subtilis* 168	NC_000964.3	Laboratory reference strain
B. mojavensis UCMB5075	NZ_CP051464.1	Mouse blood sample
B. tequilensis EA-CB0015	CP048852.1	Phyllosphere of *Musa* spp. AAA cv. Grand Naine
B. vallismortis DSM 11031	CP026362.1	DSMZ[¥] type strain

[¥] Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany. * *Bacillus* Genetic Stock Center, The Ohio State University, Ohio, U.S.A.

3.2.2. Media and culture practices

All bacterial strains were maintained on tryptic soya agar (Neogen, Lansing, Michigan, U. S. A.). and tryptic soya broth (Neogen, Lansing, Michigan, U. S. A.). Cell biomass for DNA extraction was obtained from cultures in Luria Bertani broth, containing (per L): NaCl 10 g, yeast extract 5 g, tryptone 10 g, 1000 mL distilled water, and adjusted to pH 7.5.

3.2.3. DNA extraction and PCR

Bacterial DNA was extracted from cell biomass from LB broth cultures (as specified in Section 3.2.2.) using a GeneJet Genomic DNA Purification kit (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) as per the manufacturer's instructions and using the recommended Gram-positive lysis buffer. The extracted DNA was quantified and analysed for purity using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U. S. A.). Genomic DNA was stored at -20°C until use.

The housekeeping gene sequences chosen for MLSA were: 16S ribosomal RNA subunit (16S rRNA); β-subunit RNA polymerase encoded by *rpo*B; chaperone protein (Heat shock protein 40 kDa) encoded by *dna*J; and the gyrase subunit alpha (*gyr*A). Genomic fingerprinting PCR used two approaches: Repetitive extragenic palindromic PCR (Rep-PCR) using BOX A1R primer and Randomly Amplified Polymorphic DNA (RAPD-PCR) using two primers sets: OPG-11 and OPG-16. Specific primer details and references are presented in Table 3.3.

PCRs were carried out using DreamTaq 2X Green PCR Master Mix (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.A.). The PCR protocol applied 25 µL reaction volumes containing final concentrations of: 1x DreamTaq Green PCR Master Mix; 0.3 µM of each primer; ~1 ng template DNA; and sufficient nuclease-free water (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.A.) to bring the final volume to 25 µL. The PCR protocols were applied as indicated in Table A2. using a G-Storm GS1 Thermal Cycler (G-Storm, Somerset,

U. K.). The PCR products were stored at -20 °C until use. The PCR products were separated and visualised using agarose gel electrophoresis. SeaKem LE agarose (Lonza Bioscience, Basel, Switzerland) was used at concentrations of either 1.2% w/v (Rep- and RAPD-PCRs) or 1% w/v (housekeeping genes) and amplicon band sizes were determined using a 1 kb DNA Ladder (GeneRuler, Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.). Housekeeping gene agarose gels used 1x sodium boric acid buffer [33] and fingerprinting and small-amplicon gels used 1x tris-borate-ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with CondaSafe stain (CondaLab, Madrid, Spain) as per manufacturer's instructions.

Amplicon band sizes were determined using with either 1 kb or 100 bp Plus DNA GeneRuler Ladders (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.). Gel images were captured under ultraviolet light using Syngene GeneSnap software (version 7.09) and visualised amplicons were analysed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint gel analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England) to generate dendrograms using the following parameters: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with either a 0.1% (RAPD-PCR) or 0.8% (Rep-PCR) tolerance.

Table 3.3. PCR primers for genomic fingerprinting and housekeeping gene amplification of *Bacillus subtilis sensu lato* isolates.

Protocol	Primers	Sequence (5'–3')*	PCR Annealing Temperature	Amplicon Size	Reference
RAPD-PCR	OPG-11	TGC CCG TCG T	36°C	Variable	Daffonchio *et al*. (1998)
	OPG-16	AGC GTC CTC C	36°C	Variable	
Rep-PCR	BOX A1R	CTA CGG CAA GGC GAC GCT GAC G	45°C	Variable	Versalovic *et al.* (1994)
16S rRNA	BacF	GGG AAA CCG GGG CTA ATA CCG GAT	65°C	~1187 bp	Heuer *et al.* (1997)
	R1378	CGG TGT GTA CAA GGC CCG GGA ACG			Garbeva *et al.* (2003)
dnaJ	*dna*J-F	GGG GTA GGT AAG AGC GCT TC	58°C	~788 bp	Connor *et al*. (2010)
	*dna*J-R	CGG CAT TTC GCA GTA AAT ATC			
gyrA	p-*gyr*A-f	CAG TCA GGA AAT GCG TAC GTC CTT	60°C	~892 bp	Roberts *et al.* (1994)
	p-*gyr*A-r	CAA CGT AAT GCT CCA GGC ATT GCT			
rpoB	*rpo*B-F	AGG TCA ACT AGT TCA GTA TGG ACG	48°C	~808 bp	Roberts *et al.* (1994)
	*rpo*B-R	GTC CTA CAT TGG CAA GAT CGT ATC			

* PCR primers were synthesised by Inqaba Biotec Laboratories (Pretoria, South Africa).

3.2.4. Amplicon sequencing and analysis

Amplified gene fragments were sequenced using an ABI 3500XL Genetic Analyzer (Applied Biosystems, Foster City, California, U.S.A.) at Inqaba Biotec Laboratories (Pretoria, South Africa). Sequences were visualised and edited in Chromas Lite (version 2.01) and consensus sequences generated using BioEdit (version 7.2.6.1) (Hall, 1999). Consensus nucleotide sequences were submitted to NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Zhang *et al., 2*000) for comparison to existing gene sequence data in the GenBank database and to source appropriate reference genomes. The BLASTn parameters for housekeeping gene sequences (viz., *gyrA, rpoB, dnaJ*) searched the NCBI nucleotide collection (nr/nt) database and the 16S rRNA BLASTn search used the 16S ribosomal RNA database. All searches were limited to the *Bacillus subtilis* group (Taxid: 653685), omitted uncultured/environmental material, and used the default MEGABLAST algorithm.

Phylogenies were generated for each housekeeping gene sequence dataset using MEGA X software (version 10.0.5.) (Pennsylvania State University) (Kumar *et al.*, 2018). The sequences were aligned in MEGA X using MUSCLE (Edgar, 2004). Phylogenies were generated using the Maximum Likelihood method and Kimura 2-parameter model (Kimura, 1980) and phylogeny testing applied a bootstrap method of 500 replicates. The initial trees applied to the heuristic search were obtained by Neighbour-Join and BioNJ algorithms which were applied to a pairwise distances matrix. This matrix was estimated using the Maximum Composite Likelihood approach, and then selecting the topology with the highest log likelihood value. Individual gene sequence phylogenies were also generated under the above-mentioned parameters for comparative purposes.

3.2.5. Multilocus analysis of housekeeping gene targets

The MLSA analysis was performed using the partial housekeeping gene sequences and counterpart sequences from selected reference genomes (Tables 3.1 and 3.2). The MLSA data was concatenated and phylogeny generated using MEGA X software (version 10.0.5.) (Pennsylvania State University) (Kumar *et al.*, 2018) using the previously-aligned and trimmed FASTA files.

3.2.6. Ecotype simulation

The ES analysis utilised the concatenated sequence dataset generated for MLSA to identify putative ecotypes amongst the isolate subset and reference strains using the ECOSIM2 tool (URL: https://github.com/sandain/ecosim) (Wood *et al*., 2020) (Date Accessed: 17 August 2020), which applied the Stable Ecotype Model defined by Koeppel *et al.* (2008) to define number of putative ecotypes, and rates of periodic selection and new ecotype formation.

3.3. Results

3.3.1. MLSA phylogeny of housekeeping gene targets

The MLSA phylogeny demonstrated a significant degree of variation amongst the *Bacillus* spp. isolates being evaluated (Figure 3.1). Isolates were grouped into clades that supported their presumptive identities as strains of either *B. velezensis* or *B. subtilis.* Within each of the two clades several sub-clades could be distinguished. Figure 3.1. indicates demarcations of clusters containing the plant-associated study isolates. Isolates falling within the *B. velezensis* clade were divided amongst five sub-clades namely: M-BV1 (viz., bnn282, bng199, and sqo277), M-BV2 (viz., bna81, CL-R51, CT-R23, bnd160, ccc103), M-BV3 (viz., cce146), and BV4 (viz., bng230, R16 bnd137, cce142, and pkl242), and M-BV5 (viz., sqo275); isolate sqo275 formed its own clade distinct from the *B. velezensis* reference strains and was assigned its own grouping M-BV5. It was also interesting to note that the type strain *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T} fell within the *B. velezensis* clade and shared a common node with the M-BV1 cluster. Two prominent sub-groupings, M-BS1 and M-BS2, were distinguished amongst isolates falling within the *B. subtilis* clade of related strains. The M-BS1 isolates (viz., bng215, sqo279, bnd156, and bnd134) clustered with reference strain *B. subtilis* subsp. *subtilis* strain 168, whereas BS2 isolates (viz., CT-R67, sqo271, bng221, B81 and bnd136) grouped very closely to reference strain *Bacillus* sp. JS.

A pairwise distance matrix was generated to corroborate isolate and reference strains evolutionary distances using MEGA X (Figure B2). This matrix highlighted the levels of similarities and differences within the groupings defined by the MLSA phylogeny, particularly

within a species. The *B. velezensis* isolates were more closely aligned in the phylogeny than the *B. subtilis* isolates, and this was reflected in the matrix. The *B. velezensis* sequences showed a comparatively higher degree of similarity to each other and to their reference strain counterparts than was observable amongst the *B. subtilis* isolates and reference strains.

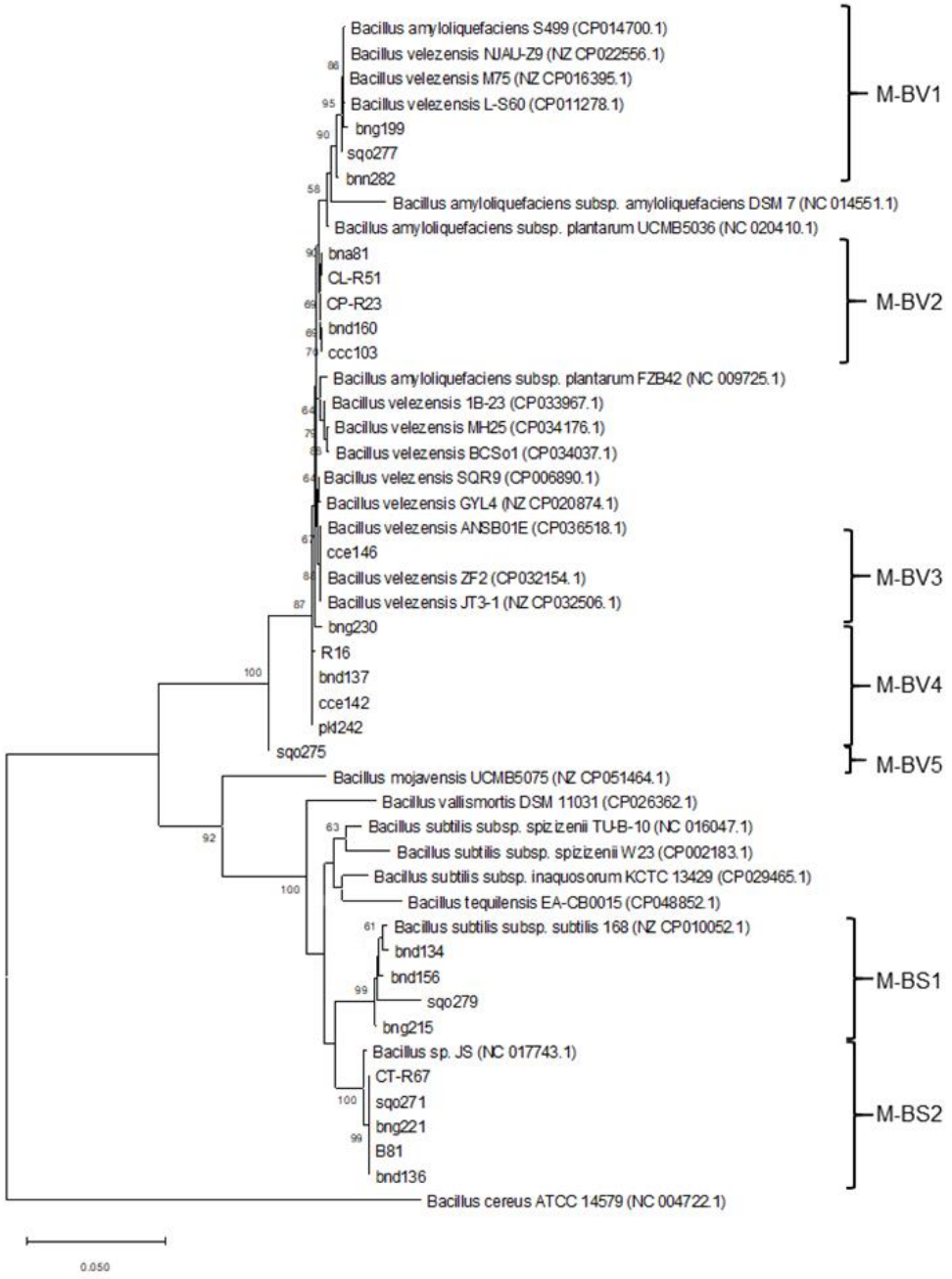


Figure 3.1. MLSA of *Bacillus subtilis sensu lato* isolates and representative reference strains. Gene sequences for reference strains were obtained from whole genome data available on the NCBI GenBank database. This 2, 839 bp phylogeny was concatenated, aligned, and generated using MEGA X (Kumar *et al*., 2018). The Maximum Likelihood method was applied, with the Kimura 2-parameter model. Initial trees applied the Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach. The tree shown is that returning the highest log value (-11713.99). The tree is drawn to scale with branch lengths measured in the number of substitutions per site, the figures next to the branches indicate the percentage trees in which the cluster appeared.

3.3.2. Putative ecotype grouping using ES analysis

The concatenated MLSA sequence data was applied to an ES analysis in order to identify putative ecotypes amongst the selected isolates and reference strains used in the study. Twelve ecotype groups were identified using the ECOSIM2 tool (Figure 3.2). The isolates accounted for eight of the ecotypes identified, with four of the ecotypes being unique and did not include any of the reference strains used in the ES analysis. With the exception of strain bng230, which grouped as a single representative of Ecotype 9, all of the ecotype groupings correlated closely with the isolate sub-groupings discerned using MLSA. For the sample set, the hillclimbing analysis projected a minimum of eight ecotypes amongst the sequences evaluated, with a rate of ecotype formation (Omega) of 0.0034 and a rate of periodic selection (Sigma) of 0.0227. The Stable Ecotype model requires that the rate of ecotype formation be lower than the rate of periodic selection in order for ecotypes to be valid within the ES algorithm, and this holds true for this dataset based on the respective omega and sigma values. The ES analysis and MLSA bear remarkable similarity in sub-clusters: Ecotype 2 comprised MLSA clade in M-BS2 strains, Ecotype 3 matched M-BS1; Ecotype 8 closely matched M-BV3, but for a deviation in the placement of bng230; Ecotype 10 comprised M-BV2; Ecotype 11 comprised M-BV1; and isolates sqo275 and cce146 were differentiated as Ecotype 7 and Ecotype 12, respectively. The MLSA and ES phylogenies show significant similarity in isolate relationships, with the exception of the grouping of reference strain *B. vallismortis* DSM 11031, which clusters in the MLSA phylogeny with the *B. subtilis* subsp. *subtilis* 168 strain grouping, and contrastingly comprises the same putative ecotype as these isolates in ES.

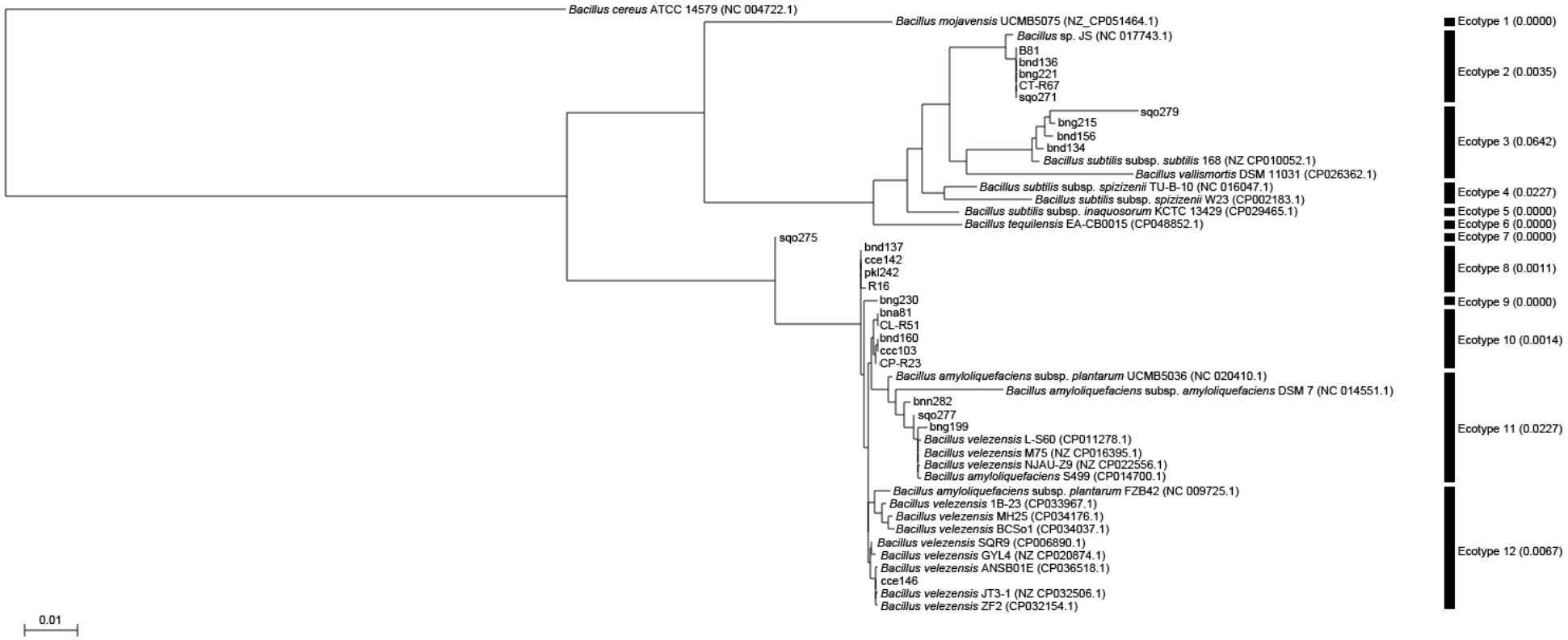


Figure 3.2. Ecotype simulation output tree for concatenated housekeeping gene sequences (16S rRNA-*gyrA*-*dnaJ*-*rpoB*) from *Bacillus subtilis sensu lato* isolates and reference strains. The ES analysis was generated using ECOSIM2 (Wood *et al*., 2020). Hillclimbing analysis extrapolated 8 ecotypes (npop), rate of ecotype formation (Omega) of 0.0034, the rate of periodic selection (Sigma) at 0.0227, with a likelihood of 0.0285. Clades identified by the Stable Ecotype algorithm to be separate ecotypes are indicated by their confidence intervals noted in brackets in the righthand side of the figure.

3.3.3. PCR fingerprinting as a potential technique to delineate ecotypes

Rep-PCR was able to distinguish several fingerprint variants amongst the isolates and reference strains evaluated (Figure B3). The dendrograms generated for these profiles allowed similar fingerprint patterns to be grouped and comparisons could be made between isolates based on their respective ecotype grouping (Figures 3.3 A and B). Minor variations in band profile were evident in some instances and this was attributed to variations in band intensity and width, which further impacted matching parameter tolerance settings.

The *B. subtilis* fingerprint profiles could be differentiated into two main clusters that corroborated the two ecotype groupings identified from the ES output. Isolate bnd134 displayed a similar banding profile to the remaining Ecotype 3 counterparts, but differed in one minor band, and all bnd134 bands were of a higher intensity. The *B. subtilis* isolates comprising Ecotype 2 (viz., B81, CT-R67, bnd136, bng221, and sqo271) showed remarkably similar banding profiles. For the *B. velezensis* isolates, the grouped fingerprint profiles (Figure 3 B) did not consistently group according to a particular ecotype. Representatives of Ecotype 10 produced distinct banding profiles that could be sub-divided into two related banding profile clusters. Isolate bng230 was the sole representative of Ecotype 9 and was found to display a unique fingerprint profile compared to the other isolates. Isolates associated with Ecotypes 7, 8, 11 and 12 exhibited fingerprint profiles that were not readily distinguished from one another. Interestingly, *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T which was assigned to Ecotype 11 produced a unique banding profile and could be differentiated from the rest of the isolates.

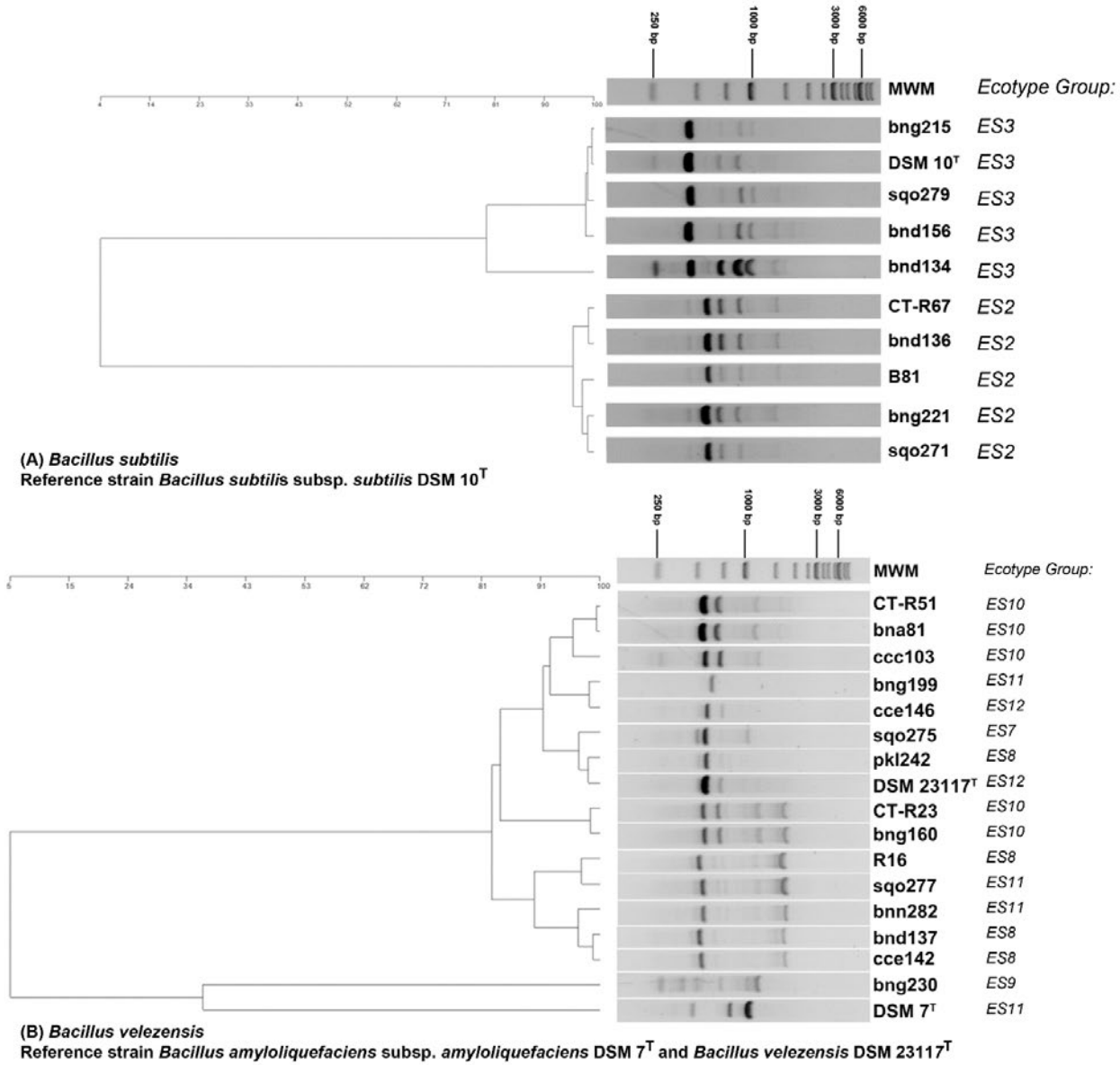


Figure 3.3. Dendrograms evaluating relationships between Rep-PCR fingerprint profiles of plant-associated *Bacillus subtilis sensu lato* isolates and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T, *B. subtilis* subsp. *subtilis* DSM 10^T, and *B. velezensis* DSM 23117^T. Fingerprint analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint dendrograms were generated using parameters of: UPGMA with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with a 0.8% tolerance. "MWM" denotes the 1 kb DNA ladder included in the agarose gels. Gel images have been inverted for clarity. Ecotype groupings for isolates and reference strains are indicated in italics.

The RAPD-PCR profiles for OPG-11 (Figure B4) and OPG-16 (Figure B5) showed different groupings between these primer sets, and in both cases showed with no matches between any of the isolates and reference strains. The OPG-11 primer set dendrogram (Figures 3.4 A and B) grouped all *B. velezensis* isolates into a single profile, which bore close similarity to that of the reference type strain *B. velezensis* DSM 23117^T, but did not differentiate at the ecotype grouping level. Reference strain *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T demonstrated a completely unique profile. The *B. subtilis* isolates comprised four profiles, with reference strain *B. subtilis subsp. subtilis* DSM 10^T being unique. The profile groupings did not correlate with the ecotype groupings. Overall, OPG-11 afforded no consistency between the profile groups and ecotype groupings amongst these isolates. The OPG-16 primer dendrogram (Figures 3.5. A and B) distinguished the reference strains *B. amyloliquefaciens* DSM 7^T, *B. subtilis* subsp. *subtilis* DSM 10^T, and *B. velezensis* DSM 23117^T from the isolate profiles. A major grouping among the *B. subtilis* isolates comprised isolates of Ecotype 2 (viz., B81, CT-R67, bnd136, bng221, and sqo271), with strains comprising Ecotype 3 (viz., bnd134, bnd156, bng215, sqo279 and *B. subtilis* subsp. *subtilis* DSM 10^T) returning profiles that were largely distinct from each other. Among the *B. velezensis* isolates a major profile grouping comprised R16, bna81, ccc103, bnd137, cce142, cce146, bnd160, pkl242, sqo277, and bnn282. This grouping, and those profiles for the remainder of the *B. velezensis* isolates, did not correlate to a specific ecotype grouping.

Ecotype groupings were not always correlated within the groupings defined on the basis of DNA fingerprinting approaches. For example, neither of isolates cce146 and sqo275 grouped with any other isolates in MLSA and ES; yet sqo275 was not differentiated in OPG-11 or Rep-PCR, but was a unique fingerprint profile in OPG-16; while isolate 146 was not differentiated in any of these. Within the *B. subtilis* isolate subset, fingerprinting methods Rep and RAPD OPG-16 consistently differentiated a fingerprint profile common to isolates B81, CT-R67, bnd136, bng221, and sqo271. These isolates all clustered with reference strain *Bacillus* sp. JS (Clade BS2) in the MLSA phylogeny, and comprised a single ecotype with this reference strain in ES analysis (Ecotype 2).

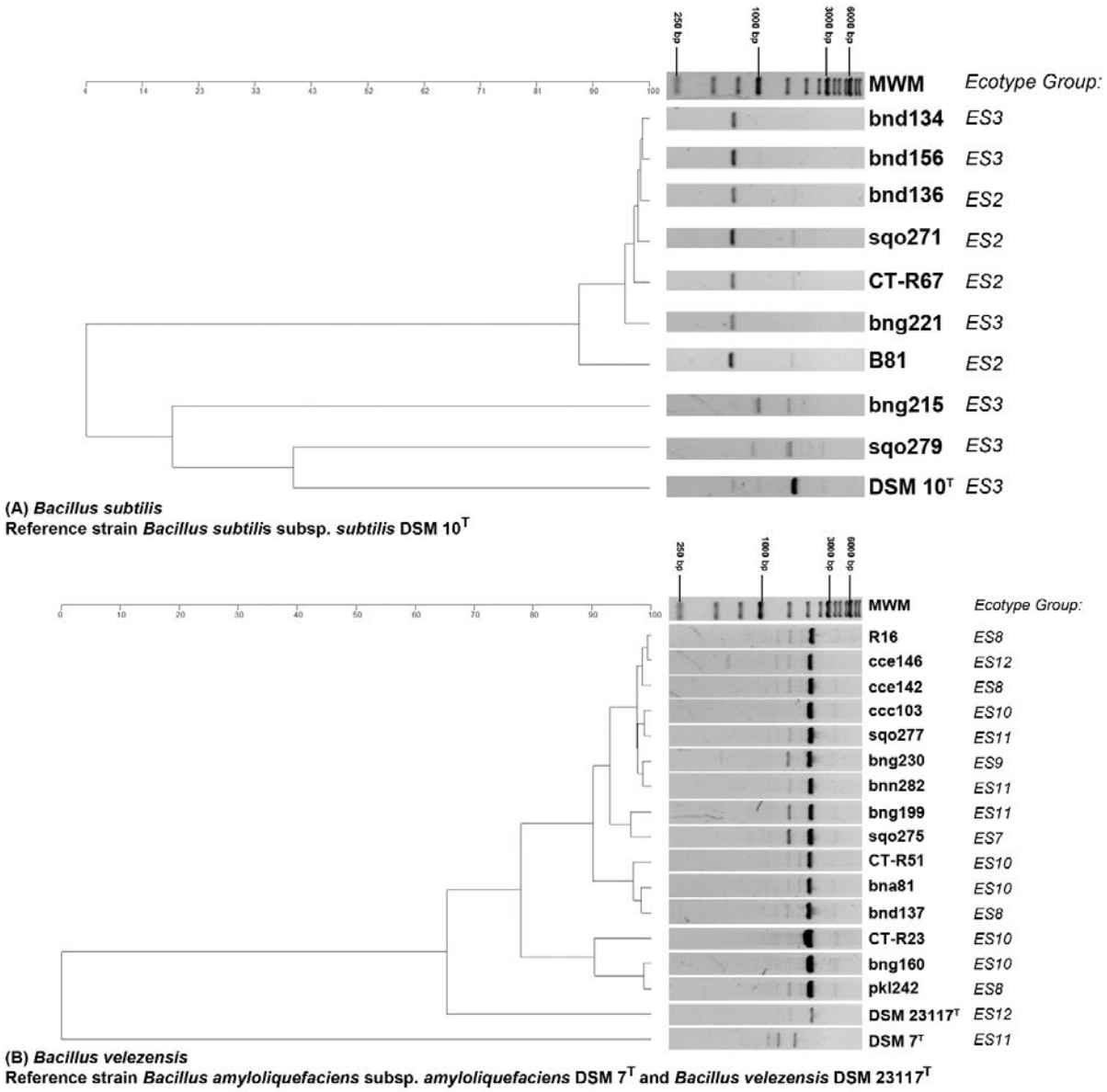


Figure 3.4. Dendrograms evaluating relationships between RAPD-PCR fingerprint profiles generated using the OPG-11 primer set for plant-associated *Bacillus subtilis sensu lato* isolates and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T, *B. subtilis* subsp. *subtilis* DSM 10^T, and *B.* velezensis DSM 23117^T. Fingerprint analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint dendrograms were generated using parameters of: UPGMA with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with a 0.1% tolerance. "MWM" denotes the 1 kb DNA ladder included in the agarose gels. Gel images have been inverted for clarity. Ecotype groupings for isolates and reference strains are indicated in italics.

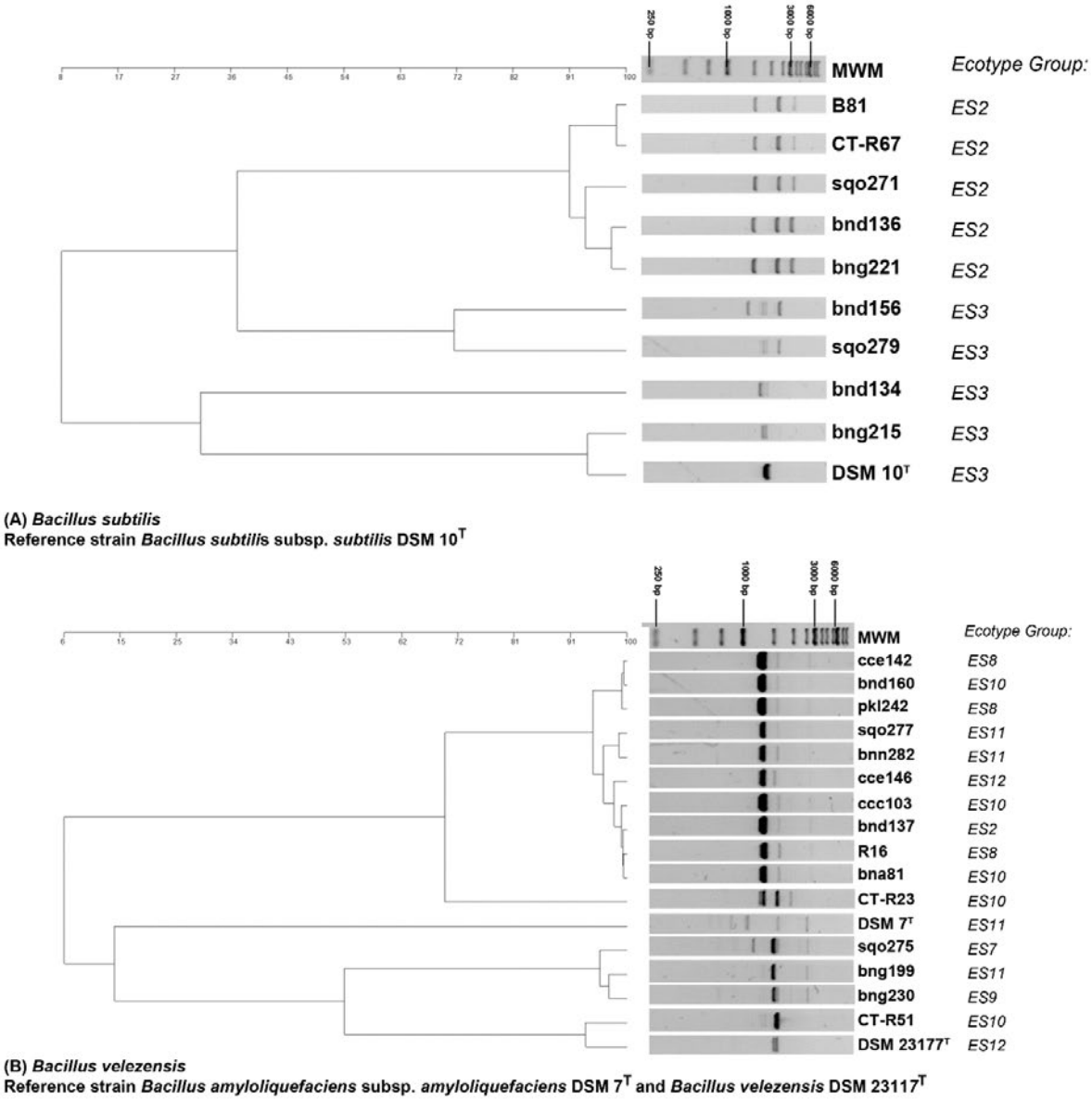


Figure 3.5. Dendrograms evaluating relationships between RAPD-PCR fingerprint profiles generated using the OPG-16 primer set for plant-associated *Bacillus subtilis sensu lato* isolates and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T, *B. subtilis* subsp. *subtilis* DSM 10^T, and *B. velezensis* DSM 23117^T. Fingerprint analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint dendrograms were generated using parameters of: UPGMA with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with a 0.1% tolerance. "MWM" denotes the 1 kb DNA ladder included in the agarose gels. Gel images have been inverted for clarity. Ecotype groupings for isolates and reference strains are indicated in italics.

3.4. Discussion

Members of the *B. subtilis sensu lato* are known to exist in association with plants and, in many instances, have been shown to positively impact plant health (Borriss, 2011, Zhang *et al.,* 2016). In the current study, MLSA and ES analysis were used to identify putative ecotypes amongst a subset of plant-associated *B. subtilis sensu lato* isolates and selected reference strains. Genomic fingerprinting approaches namely, Rep-PCR and RAPD-PCR, were then evaluated for their potential to corroborate the ecotype groupings, and as a possible means of rapid ecotyping. In the context of a rapid method, fingerprint PCR would allow the faster processing of larger numbers of isolates, thereby improving isolate throughput rates during screening. The *B. subtilis sensu lato* isolates were sourced from rhizosphere and phyllosphere samples taken from several crop plant species (viz., lettuce, tomato, various cucurbit species) cultivated in the Midlands region of the province of KwaZulu-Natal, South Africa. Isolates were selected on the basis of their ability to antagonise fungal phytopathogens *in vitro* and their ability to produce biosurfactant compounds commonly associated with biofilm regulation (Tredgold, 2015, Hunter, 2016). Isolates were identified as strains of *B. subtilis* and *B. velezensis* which displayed high levels of sequence similarity to known plant-associated reference strains. Knowledge of ecological niche specialisation in beneficial plant-associated bacterial strains is useful, as predominant ecotypes could indicate niche and habitat adaptation to localised conditions or plant type. From a plant-health-promotion perspective, it is important to select candidates that are able to establish within specific environments in order to give them the best chance of succeeding as exogenous microbial inoculants. This study represents the first instance of determining the levels of ecotype diversity amongst a subset of plant-growth-promoting *B. subtilis sensu* lato isolates from a range of crop types and localities. Additionally, to the best of our knowledge, no such studies of *B. velezensis* plant-associated ecotyping have been published, to date.

The MLSA and ES phylogenies corroborate that the South African isolates demonstrate a level of ecological variation. The level of sequence sub-clustering observed in the MLSA phylogeny may be construed as ecotype variation when comparing the South African *B. velezensis* isolates to the selected reference genomes included. Groupings that were not associated with other reference strains (e.g. M-BV2, M-BV4 and M-BV5) could be indicative of distinct ecotypes.

Ecotype variation amongst the *B. velezensis* isolates demonstrate greater levels of divergence from reference strains, in that isolates form three ecotype groups completely distinct from any of the reference strains. That isolate cce146 was the only isolate that grouped with *B. velezensis* FZB42—a well-established commercial European plant-growth-promotion product—further speaks to the localised variation of ecotypes within this species.

Other isolates were placed in groupings that included other reference strains, indicating commonality with existing ecotype groupings. Amongst the *B. subtilis* isolates, the isolate ecotypes grouped with *Bacillus* sp. JS and *B. subtilis* subsp. *subtilis* 168. *Bacillus* sp. JS is a reference strain isolated from *Miscanthus* spp. soil in Korea that was studied due to its plant-growth-promoting potential (Song *et al*., 2012); the grouping of this strain with M-BS2 isolates (selected for their biological control characteristics) gives further credence to the notion that this grouping represents a de facto plant-associated ecotype grouping. The clustering of isolates with *Bacillus* sp. JS is interesting and unexpected. In both the MLSA and ES phylogenies this *Bacillus* sp. JS cluster formed a distinct sub-clade within the *B. subtilis* reference strain group, and generate common profiles in Rep-PCR and RAPD-PCR OPG-16. This could suggest subspecies divergence of *Bacillus* sp. JS and the isolates phylogenetically grouped with it. Furthermore, that this reference strain originated in Korea and the isolates were from South Africa provides some evidence for a plant-associated ecotype which spans geographical location. Novel ecotypes of potential plant-growth-promoting *Bacillus* spp. from South Africa, and that these ecotypes comprise isolates from different crop species, suggests that there are novel ecotypes suited to the South African climate and ecological niches, and which may be applicable to a range of crop species.

MLSA was then able to define phylogenetic clades which largely corroborated those defined as putative ecotypes by the ES analysis. Both MLSA and ES methods afford a similar ease of use, with the ES analysis providing additional information as to projected rates of ecotype formation and rates of periodic selection in the supplied dataset. The selection of genomic data used for ES and MLSA is important to ensure an accurate reflection of diversity is presented. The number of gene targets applied in a MLSA depends greatly on the context and goals of that particular study (Glaeser and Kämpfer, 2015). The present study MLSA applied four gene

targets, and these individual gene sequences have been included in previous studies for their ability to differentiate between members of the *B. subtilis sensu lato* (Roberts *et al.*, 1994, Connor *et al.*, 2010, Stefanic *et al.*, 2012). A MLSA scheme typically incorporates six to ten housekeeping gene targets, though most commonly four to five are applied (Maiden, 2006, Glaeser and Kämpfer, 2015). To this end, we consider the sequence dataset applied in this study to be sufficient for the goal of defining *B. subtilis sensu lato* ecotypes.

The phylogenetic clustering observed in housekeeping gene sequence MLSA provides some basis on which to differentiate ecotype populations (Koeppel *et al.*, 2008, Dogsa *et al.*, 2014), but this approach lacks the acknowledgement of such factors as genetic drift and periodic selection which are in play in ecotype populations. The ES algorithm is more likely to identify ecotypes under the Stable Ecotype model, where ecotype undergoes multiple periodic selection events within its lifetime (Koeppel *et al.*, 2008). This allows for increase in gene sequence divergence between ecotypes, with diversity being purged within—but not between—ecotypes; which in turn results in clades corresponding between ecotypes and sequence divergence (Koeppel *et al.*, 2008). The ES analysis using the concatenated MLSA sequence data determined that the isolates varied at a level expected of ecotypes, in that the omega and sigma value relationship corroborated the Stable Ecotype model.

In bacterial pathogen epidemiology studies, fingerprint profiling has been shown to reflect the levels of diversity achieved using multilocus gene sequence approaches (Sachse *et al.*, 2014, Johnson-Mackinnon *et al.*, 2019). For ecotyping purposes DNA fingerprinting approaches have proven useful in discriminating between closely-related bacterial strains, and in particular RAPD-PCR, have demonstrated the potential to define subpopulation diversity and ecotypes amongst *Bacillus* spp. (Daffonchio *et al.*, 1998, Istock *et al.*, 2001, Sikorski and Nevo, 2005, Kwon *et al.*, 2009). The approach does not require the amplification and sequencing of gene targets, and would be convenient as a high-throughput ecotyping method for the evaluation of large subsets of isolates. The fingerprinting results for this study were varied. Rep-PCR proved most useful in that the BOX-A1R primer used was able to differentiate strains at the species level, and in some instance, was able to differentiate certain ecotype groupings (e.g. Ecotypes 2, 3, 9 and 10). This was more evident for the *B. subtilis*-related strains which were clearly

differentiated into the sub-clades established when MLSA was applied. The *B. velezensis* ecotype clades were less clear when Rep-PCR was applied and this may be attributed to variations in parts of the genome not linked to the housekeeping genes applied in this study. Fingerprinting using the RAPD OPG-16 primer allowed Ecotypes 2 and 3 to be distinguished but were unable to separate the *B. velezensis* ecotype clusters. Primer OPG-11 fingerprint profiles yielded the lowest levels of differentiation amongst strains and was therefore, considered to be of limited value for identifying ecotypes. However, the OPG-11 primer was able to distinguish a single profile in common to all the *B. velezensis* isolates. This profile varied in the intensity of some of the minor bands, and was closely related to that of the *B. velezensis* DSM 23117^{T} reference strain (=FZB42), but distinct to the *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T}, which suggests some application potential for this fingerprinting PCR primer as a method for the presumptive identification of *B. velezensis* strains. Furthermore, the OPG-11 *B. velezensis* profile also demonstrated a common band the approximate size of 2000 bp (Figure 4 B) which may hold some future potential as a marker for plant-associated *B. velezensis.*

Fingerprinting is by its nature highly sensitive, and targets variations across the whole genome, rather than the more restricted view afforded by housekeeping sequence data. This may explain why fingerprint profiles were not limited to single MLSA and ES clades. These fingerprinting approaches also generated the multiple band amplicons by different means: Rep-PCR targeted repetitive palindromic regions throughout the genome using a single BOXA1 primer set, while RAPD-PCR used short random primers that anneal at sites throughout the genome to generate strain specific fingerprint profiles. The three fingerprinting primer sets used in this study are merely a portion of those available, and, given the discrimination capacity of fingerprinting PCR, it could be argued that expanding the range of primers evaluated may find one or more suitable high-throughput ecotype discrimination. Moreover, other parameters related to PCR may be further optimised. There is also room to expand the multilocus sequence dataset in an effort endeavour to more favourably align the outcomes of these two methods (Fingerprinting and multilocus analysis).

3.4.1. Conclusions

Bacillus spp. are known to become adapted to certain regions, soil types, and environments, to the point of being defined as ecotypes (Reva *et al.*, 2004, Sikorski and Nevo, 2005, Connor *et al.*, 2010, Liu *et al.*, 2013, Swiecicka *et al.*, 2013, Sachse *et al.*, 2014, Zhang *et al.*, 2016). The *Bacillus* sp. JS clade of isolates (Ecotype 2) speaks to this, with the isolates included in this clade originating from multiple crop species and locations. Throughout the data generated by this study it is apparent that there exists some degree of locality-specific variation in some of the isolates when compared to the representative *Bacillus* spp. genomes included. Local ecological adaptation of beneficial *Bacillus subtilis sensu lato* has implications in the agricultural context, in that the populations present in soils and in association with crop plants from a particular region may become ecologically distinct from those found elsewhere, and, in this case, may have become niche-adapted to the localised conditions (e.g. climate and soil type) present in KwaZulu-Natal. The fact that the plant-associated *B. velezensis* isolates demonstrated a range of diversity and formed several distinct putative ecotype groupings is also noteworthy.

With the global agricultural spotlight focussed on sustainable agriculture and responsible soil stewardship, recognising the presence and diversity of native beneficial bacterial species—like *Bacillus* spp.—holds promise for identifying and selecting promising candidates as biofertilizers and biocontrol agents capable of competing and persisting within the prevailing conditions of the local environment. The ecological variation amongst the subset of plant-associated *B. subtilis sensu lato* isolates used in this study, and the initial selection bias for biocontrol traits, allows some postulation as to the isolates' predisposition to the unique plant-associated conditions prevailing in KwaZulu-Natal, which in turn may carry implications for locally-derived microbial biofertilizers and biocontrol agents. Appreciating local microbial population variation could greatly contribute to our understanding of South African plant-microbe ecology and beneficial *Bacillus* species extant in local soils. Furthermore, we may better comprehend the ability of such organisms to exist as targeted biocontrol agents and biofertilizers already adapted to specific agricultural niches, crop species, and cultivation conditions. A future avenue for this research would be to verify whether apparent "locally-adapted" strains could provide superior plant health promotion activities under field conditions.

3.5. References

Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L. & Olmedo, G. (2010) Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics* **11**: 332, 1–17. 10.1186/1471-2164-11-332

Borriss, R. (2011) Use of plant-associated *Bacillus* strains as biofertilizers and biocontrol agents in agriculture. *Bacteria in Agrobiology: Plant Growth Responses* (Maheshwari, D. K., ed.) pp. 41–76. Springer-Verlag, Berlin, Germany. 10.1007/978-3-642-20332-9_3

Brody, J. R. & Kern, S. E. (2004) Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *BioTechniques* **36**: 2, 214–216. 10.2144/04362BM02

Cawoy, H., Bettiol, W., Fickers, P. & Ongena, M. (2011) *Bacillus*-based biological control of plant diseases. *Pesticides in the Modern World: Pesticides Use and Management* (Stoytcheva, M., ed.) pp. 273–302. InTech. 10.5772/17184

Chen, X. H., Koumoutsi, A., Scholz, R. & Borriss, R. (2009a) More than anticipated: Production of antibiotics and other secondary metabolites by *Bacillus amyloliquefaciens* FZB42. *Journal of Molecular Microbiology and Biotechnology* **16**: 1–2, 14–24. 10.1159/000142891

Chen, X. H., Koumoutsi, A., Scholz, R., Schneider, K., Vater, J., Süssmuth, R., Piel, J. & Borriss, R. (2009b) Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *Journal of Biotechnology* **140**: 27–37. 10.1016/j.jbiotec.2008.10.011

Cohan, F. M. (2005) Periodic selection and ecological diversity in bacteria. *Selective Sweep* (Nurminsky, D., ed.) pp. 78–93. Springer US, Boston, MA. 10.1007/0-387-27651-3_7

Cohan, F. M. (2006) Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **361**: 1475, 1985–1996. 10.1098/rstb.2006.1918

Cohan, F. M. & Koeppel, A. F. (2008) The origins of ecological diversity in prokaryotes. *Current Biology* **18**: 21, 1024–1034. 10.1016/j.cub.2008.09.014

Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koeppel, A. F., Burger, A., Cole, S. G., Perry, E. B., Krizanc, D., Field, N. C., Slaton, M. & Cohan, F. M. (2010) Ecology of speciation in the genus *Bacillus*. *Applied and Environmental Microbiology* **76**: 5, 1349–1358. 10.1128/aem.01988-09

Daffonchio, D., Borin, S., Frova, G., Manachini, P. L. & Sorlini, C. (1998) PCR fingerprinting of whole genomes: the spacers between the 16S and 23S rRNA genes and of intergenic tRNA gene regions reveal a different intraspecific genomic variability of *Bacillus cereus* and *Bacillus licheniformis*. *lnternational Journal of Systematic Bacteriology* **48**: 107–116. 10.1099/00207713-48-1-107

Dogsa, I., Oslizlo, A., Stefanic, P. & Mandić-Mulec, I. (2014) Social interactions and biofilm formation in *Bacillus subtilis*. *Food Technology and Biotechnology* **52**: 2, 149–157. 10.1128/JB.01290-08

Dunlap, C. A. (2019) Taxonomy of registered *Bacillus* spp. strains used as plant pathogen antagonists. *Biological Control* **134**: 82–86. https://doi.org/10.1016/j.biocontrol.2019.04.011

Edgar, R. C. (2004) MUSCLE: Multiple alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 5, 1792–1797. 10.1093/nar/gkh340

Fajardo-Cavazos, P., Maughan, H. & Nicholson, W. L. (2014) Evolution in the *Bacillaceae*. *Microbiology Spectrum* **2**: 5, 1–32. 10.1128/microbiolspec.TBS-0020-2014

Fan, B., Borriss, R., Bleiss, W. & Wu, X. (2012) Gram-positive rhizobacterium *Bacillus amyloliquefaciens* FZB42 colonizes three types of plants in different patterns. *The Journal of Microbiology* **50**: 1, 38–44. 10.1007/s12275-012-1439-4

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. (2009) The Bacterial Species Challenge: Making Sense of Genetic and Ecological Diversity. *Science* **323**: 741–746. 10.1126/science.1159388.

Fritze, D. (2004) Taxonomy of the genus *Bacillus* and related genera: The aerobic endospore-forming bacteria. *Phytopathology* **94**: 11, 1245–1248. https://doi.org/10.1094/PHYTO.2004.94.11.1245

Garbeva, P., van Veen, J. A. & van Elsas, J. D. (2003) Predominant *Bacillus* spp. in agricultural soil under different management regimes detected via PCR-DGGE. *Microbial Ecology* **25**: 3, 302–316. 10.1007/s00248-002-2034-8

Glaeser, S. P. & Kämpfer, P. (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology* **38**: 4, 237–245. http://dx.doi.org/10.1016/j.syapm.2015.03.007

Hall, T. A. (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98. 10.14601/Phytopathol_Mediterr-14998u1.29

Heuer, H., Krsek, M., Baker, P., Smalla, K. & Wellington, E. M. (1997) Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Applied and Environmental Microbiology* **63**: 3233–3241. 10.1128/AEM.63.8.3233-3241.1997

Hunter, C. H. (2016) Screening of aerobic endospore-forming bacterial isolates as candidate biocontrol agents against *Rhizoctonia solani*. Ph.D. Thesis Thesis, University of KwaZulu-Natal, Pietermaritzburg, KwaZulu-Natal, South Africa.

Ibarz Pavon, A. B. & Maiden, M. C. (2009) Multilocus sequence typing. *Molecular Epidemiology of Microorganisms Methods in Molecular Biology (Methods and Protocols)* Vol. 551 (Caugant, D., ed.) pp. 129–140. Humana Press, Totowa, NJ, USA. 10.1007/978-1-60327-999-4_11

Istock, C. A., Ferguson, N., Istock, N. L. & Duncan, K. E. (2001) Geographical diversity of genomic lineages in *Bacillus subtilis* (Ehrenberg) Cohn *sensu lato*. *Organisms Diversity & Evolution* **1**: 3, 179–191. https://doi.org/10.1078/1439-6092-00017

Johnson-Mackinnon, J. C., Crosbie, P. B. B., Karlsbakk, E., Marcos-Lopez, M., Paley, R., Nowak, B. F. & Bridle, A. R. (2019) Multilocus sequence typing (MLST) and random polymorphic DNA (RAPD) comparisons of geographic isolates of *Neoparamoeba perurans*, the causative agent of amoebic gill disease. *Pathogens* **8**: 4, 244–262. 10.3390/pathogens8040244

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 2, 111–120. 10.1007/bf01731581

Koeppel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., Rooney, A. P., Brambilla, E., Connor, N., Ratcliff, R. M., Nevo, E. & Cohan, F. M. (2008) Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 7, 2504–2509. 10.1073/pnas.0712205105

Kopac, S., Wang, Z., Wiedenbeck, J., Sherry, J., Wu, M. & Cohan, F. M. (2014) Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. *Applied and Environmental Microbiology* **80**: 16, 4842–4853. 10.1128/aem.00576-14

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 6, 1547–1549. 10.1093/molbev/msy096

Kwon, G.-H., Lee, H.-A., Park, J.-Y., S., K. J., Lim, J., Park, C.-S., Kwon, D.-Y., Kim, Y.-S. & Kim, J. H. (2009) Development of a RAPD-PCR method for identification of *Bacillus* species isolated from Cheonggukjang. *International Journal of Food Microbiology* **129**: 282–287. 10.1016/j.ijfoodmicro.2008.12.013

Lan, R. & Reeves, P. R. (2000) Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends in Microbiology* **8**: 9, 396–401. 10.1016/S0966-842X(00)01791-1

Lassalle, F., Muller, D. & Nesme, X. (2015) Ecological speciation in bacteria: Reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. *Research in Microbiology* **166**: 10, 729–741. http://dx.doi.org/10.1016/j.resmic.2015.06.008

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G. & Shao, Z. (2013) Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. *PLoS One* **8**: 11, e80097. 10.1371/journal.pone.0080097

Maiden, M. C. (2006) Multilocus sequence typing of bacteria. *Annual Review of Microbiology* **60**: 561–588. 10.1146/annurev.micro.59.030804.121325

Mandić-Mulec, I., Stefanic, P. & van Elsas, J. D. (2015) Ecology of *Bacillaceae*. *Microbiology Spectrum* **3**: 1, 1–23. 10.1128/microbiolspec.TBS-0017-2013

Reva, O. N., Dixelius, C., Meijer, J. & Priest, F. G. (2004) Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *FEMS Microbiology Ecology* **48**: 249–259. 10.1016/j.femsec.2004.02.003

Reva, O. N., Swanevelder, D. Z. H., Mwita, L. A., Mwakilili, A. D., Muzondiwa, D., Joubert, M., Chan, W. Y., Lutz, S., Ahrens, C. H., Avdeeva, L. V., Kharkhota, M. A., Tibuhwa, D., Lyantagaye, S., Vater, J., Borriss, R. & Meijer, J. (2019) Genetic, epigenetic and phenotypic diversity of four *Bacillus velezensis* strains used for plant protection or as probiotics. *Frontiers in Microbiology* **10**: 2610, 1–25. 10.3389/fmicb.2019.02610

Roberts, M. S., Nakamura, L. K. & Cohan, F. M. (1994) *Bacillus mojavensis* sp. nov., distinguishable from *Bacillus subtilis* by sexual isolation, divergence in DNA sequence, and differences in fatty acid composition. *International Journal of Systematic Bacteriology* **44**: 2, 256–264. 10.1099/00207713-44-2-256

Sachse, S., Bresan, S., Erhard, M., Edel, B., Pfister, W., Saupe, A. & Rödel, J. (2014) Comparison of multilocus sequence typing, RAPD, and MALDI-TOF mass spectrometry for typing of β-lactam-resistant *Klebsiella pneumoniae* strains. *Diagnostic Microbiology and Infectious Disease* **80**: 4, 267–271. 10.1016/j.diagmicrobio.2014.09.005

Schleifer, K.-H. (2009) The Firmicutes. *Bergey's Manual of Systematic Bacteriology* Vol. 3 (Parte, A. C., Whitman, W. B., De Vos, P., Garrity, G. M., Jones, D., Krieg, N. R., Ludwig, W., Rainey, F. A. & Schleifer, K.-H., eds.), pp. 19–128. Springer, New York, NY. 10.1002/9781118960608

Sikorski, J. & Nevo, E. (2005) Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 44, 15924–15929. 10.1073/pnas.0507944102

Song, J. Y., Kim, H. A., Kim, J.-S., Kim, S.-Y., Jeong, H., Kang, S. G., Kim, B. K., Kwon, S.-K., Lee, C. H., Yu, D. S., Kim, B. S., Kim, S.-H., Kwon, S. Y. & Kim, J. F. (2012) Genome sequence of the plant growth-promoting rhizobacterium *Bacillus* sp. strain JS. *Journal of Bacteriology* **194**: 14, 3760–3761. 10.1128/jb.00676-12

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F. M. & Mandić-Mulec, I. (2012) The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. *Environmental Microbiology* **14**: 6, 1378–1389. 10.1111/j.1462-2920.2012.02717.x

Swiecicka, I., Bartoszewicz, M., Kasulyte-Creasey, D., Drewnowska, J. M., Murawska, E., Yernazarova, A., Lukaszuk, E. & Mahillon, J. (2013) Diversity of thermal ecotypes and potential pathotypes of *Bacillus thuringiensis* soil isolates. *FEMS Microbiology Ecology* **85**: 2, 262–272. 10.1111/1574-6941.12116

Tredgold, H. R. (2015) Screening for aerobic endospore-forming bacteria as biocontrol agents for powdery mildew disease of cucurbits. M.Sc. Thesis, University of KwaZulu-Natal, Pietermaritzburg, KwaZulu-Natal, South Africa.

Versalovic, J., Schneider, M., De Bruijn, F. & Lupski, J. R. (1994) Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. *Methods in Molecular and Cell Biology* **5**: 25–40. https://doi.org/10.1007/978-1-4615-6369-3_34

Ward, B. B. (2002) How many species of prokaryotes are there? *Proceedings of the National Academy of Sciences of the United States of America* **99**: 16, 10234–10236. 10.1073/pnas.162359199

Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kuhl, M. & Grossman, A. (2007) Genomics, environmental genomics and the issue of microbial species. *Heredity* **100**: 2, 207–219. 10.1038/sj.hdy.6801011

Wiedenbeck, J. & Cohan, F. M. (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* **35**: 5, 957–976. 10.1111/j.1574-6976.2011.00292.x

Wood, J. M., Becraft, E. D., Krizanc, D., Cohan, F. M. & Ward, D. M. (2020) Ecotype Simulation 2: An improved algorithm for efficiently demarcating microbial species from large sequence datasets. *bioRxiv* 2020.2002.2010.940734. 10.1101/2020.02.10.940734

Zhang, N., Yang, D., Kendall, J. R. A., Borriss, R., Druzhinina, I. S., Kubicek, C. P., Shen, Q. & Zhang, R. (2016) Comparative genomic analysis of *Bacillus amyloliquefaciens* and *Bacillus subtilis* reveals evolutional traits for adaptation to plant-associated habitats. *Frontiers in Microbiology* **7**: 2039, 1–14. 10.3389/fmicb.2016.02039

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**:1/2, 203–214. 10.1089/10665270050081478

CHAPTER FOUR

ELUCIDATING PHEROTYPES AMONGST PLANT-ASSOCIATED *BACILLUS SUBTILIS SENSU LATO* ISOLATES

4.1. Introduction

The ability of bacterial populations to cooperate and form biofilms is reliant upon quorum sensing (Camilli and Bassler, 2006). Quorum sensing is a density-dependent form of bacterial communication enabled by the release of diffusible self-produced signals—known as pheromones or autoinducers—into the surrounding environment to facilitate coordination of gene expression and population behaviour (Joint *et al.*, 2007, Williams, 2007). Quorum sensing signals may be highly specific, which aids the producing population in establishment in the habitat and niche advancement, while excluding surrounding competitors and cheaters (Williams, 2007). One of the quorum-sensing mechanisms used by the *Bacillus subtilis sensu lato* group employs the ComX autoinducer peptide (Solomon *et al.*, 1995, Solomon *et al.*, 1996). ComX is part of the *comQXPA* quorum-sensing cassette responsible for the synthesis, post-translational modification, release, and detection of the mature ComX peptide (Tortosa *et al.*, 2001, Bacon Schneider *et al.*, 2002, Ansaldi and Dubnau, 2004). This is a major system in the *Bacillus subtilis sensu lato*, and ComX-mediated quorum sensing plays roles in a range of processes, including biofilm development and subpopulation differentiation, sporulation and competence development, and surfactin production (Solomon *et al.*, 1996, Tortosa and Dubnau, 1999, Tortosa *et al.*, 2001, Hamoen *et al.*, 2003). ComX contains a post-translational modification in the form of an isoprenoid side chain on a conserved tryptophan residue, with the side chain length varying in length between variants (Okada, 2011). The range of polymorphisms in ComX result in population-specific communication groupings known as pherotypes (Tran *et al.*, 2000, Ansaldi *et al.*, 2002).

Stefanic and Mandić-Mulec (2009) suggested that diversification into distinct communication groupings represents a functional diversification affording ecological adaptation. The evidence

for communication divergence by ComX variation and associated ecological specialisation has been described in other studies of soil- and rhizoplane-dwelling *B. subtilis sensu lato* species (Stefanic and Mandić-Mulec, 2009, Stefanic *et al.*, 2012, Oslizlo *et al.*, 2015). These studies have proposed a link between the variability in ComX pheromones and plant- and soil-associated diversification that has been observed in *B. subtilis sensu lato* strains. Ecotype populations in *B. subtilis sensu lato* have been described in several studies (Reva *et al.*, 2004, Sikorski and Nevo, 2005, Connor *et al.*, 2010, Koeppel *et al.*, 2013). Yet, comparatively few studies have examined pherotype divergence within populations of these species, with no studies addressing pherotype variation in plant-beneficial *B. velezensis* strains. The present study aims to investigate pherotype variants amongst a subset of plant-associated *B. subtilis sensu lato* isolates of biocontrol potential—namely *B. subtilis* and *B. velezensis*—using sequence data, PCR-RFLP, and a reporter gene assay.

4.2. Materials and Methods

4.2.1. Bacterial strains

The plant-associated strains used in this study were related to the *B. subtilis sensu lato* group, and comprised *B. subtilis* and *B. velezensis* species (Table 4.1). Also included were type strains *B. subtilis* subsp. *subtilis* DSM 10^T, *B. velezensis* DSM 23117^T, and *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T (Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany). A subset of seven *Bacillus* spp. pherotype reference strains were gifted to this study (Dr. David Dubnau, Public Health Research Institute, New Jersey Medical School, Rutgers University, U. S. A.) for use in reporter gene assay and as pherotype reference strains (Table 4.2.).

Table 4.1. *Bacillus subtilis sensu lato* isolates evaluated in this study. Isolates originated from phyllosphere and rhizosphere samples from various crop plant species in the KwaZulu-Natal province, South Africa).

Isolate	Species	Plant Species and Locality	Reference
bna81	*B. velezensis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
ccc103	*B. velezensis*	Cucumber leaf, KwaZulu-Natal	Tredgold (2015)
bnd137	*B. velezensis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
cce142	*B. velezensis*	Cucumber leaf, KwaZulu-Natal	Tredgold (2015)
cce146	*B. velezensis*	Cucumber leaf, KwaZulu-Natal	Tredgold (2015)
bnd160	*B. velezensis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng199	*B. velezensis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng230	*B. velezensis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
pkl242	*B. velezensis*	Pumpkin leaf, KwaZulu-Natal	Tredgold (2015)
sqo275	*B. velezensis*	Squash leaf, KwaZulu-Natal	Tredgold (2015)
sqo277	*B. velezensis*	Squash leaf, KwaZulu-Natal	Tredgold (2015)
bnn282	*B. velezensis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
R16	*B. velezensis*	Lettuce root, KwaZulu-Natal (Medium: soil)	Hunter (2016)
CP-R23	*B. velezensis*	Capsicum root, KwaZulu-Natal (Medium: Composted pine bark)	Hunter (2016)
CL-R51	*B. velezensis*	Lettuce root, KwaZulu-Natal (Medium: Composted pine bark)	Hunter (2016)
bnd134	*B. subtilis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bnd136	*B. subtilis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bnd156	*B. subtilis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng215	*B. subtilis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
bng221	*B. subtilis*	Butternut leaf, KwaZulu-Natal	Tredgold (2015)
sqo271	*B. subtilis*	Squash leaf, KwaZulu-Natal	Tredgold (2015)
sqo279	*B. subtilis*	Squash leaf, KwaZulu-Natal	Tredgold (2015)
CT-R67	*B. subtilis*	Tomato root, KwaZulu-Natal (Medium: Composted pine bark)	Hunter (2016)
B81	*B. subtilis*	Pumpkin roots, KwaZulu-Natal (Medium: soil)	Hunter (2016)

Table 4.2. Attributes of pherotype *Bacillus* spp. pherotype strains.

Pherotype	Strain	ComX Attribute*	Antibiotic Resistance Profile
168	BD 630	Type strain	-
	BD 2833	Producer	Tetracycline
	BD 2876	Tester	Tetracyline
RO-B-2	BD 2917	Type strain	-
	BD 2936	Producer	Chloramphenicol, tetracyline
	BD 2983	Tester	Chloramphenicol, tetracycline, spectinomycin
RO-C-2	BD 2918	Type strain	-
	BD 2937	Producer	Chloramphenicol, tetracyline
	BD 2963	Tester	Tetracyline, erythromycin
RO-E-2	BD 2925	Type strain	-
	BD 2940	Producer	Chloramphenicol, tetracyline
	BD 3020	Tester	Chloramphenicol, tetracycline, spectinomycin
RO-FF-1	BD 2922	Type strain	-
	BD 2939	Producer	Chloramphenicol, tetracyline
	BD 2992	Tester	Chloramphenicol, tetracycline, spectinomycin
RO-H-1	BD 2977	Type strain	-
	BD 2913	Producer	Tetracyline, erythromycin
	BD 2962	Tester	Tetracyline, erythromycin, spectinomycin
RS-D-2	BD 2924	Type strain	-
	BD 2949	Producer	Chloramphenicol, tetracyline
	BD 3019	Tester	Chloramphenicol, tetracycline, spectinomycin

* Pherotype strains were previously engineered into "Producer" strains able to synthesise ComX but unable to detect it, or as "Tester" strains unable to synthesise ComX capable of responding to the pheromone (Ansaldi *et al.*, 2002; Tortosa *et al.*, 2001), strains labelled as "Type strain" are unmodified.

4.2.2. Media and culture practices

All isolates and strains were maintained in long-term storage as 20 % (v/v) glycerol stocks of 48-hour-old culture (28 °C, shaken at 150 rpm) grown in tryptic soya broth (TSB) (Neogen, Lansing, Michigan, U. S. A.). Bacterial strains were maintained on tryptic soya agar (TSA) (Neogen, Lansing, Michigan, U. S. A.). The culture medium used for ComX pheromone assay and conditioned medium was a modified S7 minimal medium (Bacon-Schneider *et al.*, 2002; Vasantha and Freese, 1980) comprised of the following: K_2HPO_4 5 mM (pH 7.0), $(NH_4)_2SO_4$ 10 mM, morpholinepropanesulfonic acid (MOPS) (pH 7.0 using 10 M KOH), 50 mM, $MgCl_2.6H_2O$ 2 mM, $CaCl_2.2H_2O$ 0.7 mM, $MnCl_2.4H_2O$ 0.05 mM, $FeCl_3.6H_2O$ 5 µM, and $ZnCl_2$ 1 µM. Glucose was autoclaved separately, and added at a final concentration of glucose 1% (w/v). Filter-sterilised (0.22 micron) additions of vitamins, amino acids, and appropriate antibiotics were added after autoclaving the basal medium. Remaining components were added

to achieve the following final concentrations: thiamine 2 µM, tryptophan and phenylalanine 40 µg/mL, methionine 25 µg/mL, threonine 120 µg/mL, sodium glutamate (pH 7.0) 0.1 % (w/v) (Bacon-Schneider *et al.*, 2002). Antibiotic amendments for the producer and tester pherotype strains (Table 4.2.) were added the following concentrations: tetracycline 20 µg/mL, spectinomycin 100 µg/mL, erythromycin 5 µg/mL, chloramphenicol 5 µg/mL, neomycin 5 µg/mL, and rifampin 5 µg/mL (Melford Laboratories Ltd., Ipswich, U.K.).

The prepared S7 medium was inoculated from starter cultures grown in TSB (Neogen, Lansing, Michigan, U. S. A.) with appropriate antibiotic amendments. Starter cultures comprised a single-colony-inoculation into 10 mL TSB which was grown overnight at 30 °C at 150 rpm in a shaking incubator. Aliquots (1 mL) of starter culture were centrifuged (9, 503 *x g* for 5 minutes) and the cell pellet resuspended in an equal amount sterile saline (0.85 % NaCl) to allow spectrophotometric measurement of the optical density (OD) (A_{600}). This value was then used to determine appropriate amounts of supernatant-free cell pellet to add to the S7 medium to achieve the desired OD.

4.2.3. DNA extraction and quantification

Cell biomass for DNA extraction was prepared as single bacterial colonies (grown overnight on TSA at 30 °C) inoculated into 10 mL aliquots of sterile Luria Bertani broth, containing (per L): NaCl 10 g, yeast extract 5 g, tryptone 10 g, 1000 mL distilled water, and adjusted to pH 7.5. Broths were inoculated with a single colony and were incubated overnight at 30 °C at 150 rpm in a shaking incubator before genomic DNA extraction. Bacterial DNA was extracted using a GeneJet Genomic DNA Purification kit (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) as per the manufacturer's instructions and using the recommended Gram-positive lysis buffer. The extracted DNA was quantified and analysed for purity using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.). Genomic DNA was stored at -20 °C until use.

4.2.4. PCR amplification of *comQXP* gene loci

The *comQXPA* quorum sensing loci for *B. subtilis* isolates was amplified using primers obtained from Stefanic and Mandic-Mulec (2009), with the forward primer Uni-comQ1 5'-GGGAGGGGGGAAGTCGTTATTG-3' and reverse primer P1 5'-AAGAACCGAATCGTGGAGATCGCG-3'. The *comQXPA* quorum sensing loci for *B. velezensis* isolates necessitated the design of a novel set of suitable primers (Chapter 2): Forward primer ComBamyF 5'-GACATCTGCTCATTCATCTGATGAAG-3' and reverse primer ComBamyR 5'-CAGAGAAACAACCGACTCATTACG-3'. The ComBamy *comQXP* primer set was derived for the *B. velezensis* species using the online tool PrimerBLAST (URL: https://www.ncbi.nlm.nih.gov/tools/primer-blast) (Ye *et al.*, 2012) applying reference genomes *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7 (NC_014551.1) and *B. velezensis* FZB42 (NC_009725.2). The PCR protocols used a G-Storm GS1 Thermal Cycler (G-Storm, Somerset, U. K.). Both primer sets yielded expected fragments ~3 kb in length. Respective PCR protocols were carried out using a G-Storm GS1 Thermal Cycler (G-Storm, Somerset, U.K.). PCR products were stored at -20 °C until use.

Amplification of the *B. subtilis* isolates' genomic DNA using a KAPA HiFi HotStart PCR kit (KAPA Biosystems, Wilmington, Massachusetts, U.S.A.). The PCR protocol applied 50 µl reactions volumes containing final concentrations of: 1X KAPA HiFi Fidelity Buffer, 0.3mM each deoxynucleoside triphosphate (dNTP), 0.3 µM each of forward and reverse primers, ~1 ng template DNA, 0.5 U/reaction KAPA HiFi Hotstart DNA polymerase, with sufficient nuclease-free water (Promega, Madison, U.S.A.) to bring the final volume to 50 µl. The PCR protocol applied initial denaturation at 95 °C for 4 minutes; and 30 cycles of: denaturation at 98 °C for 20 seconds; annealing at 56 °C for 15 seconds; and extension for 1 minute at 72 °C; with a final extension of 3 minutes at 72 °C.

Amplification of the *B. velezensis* quorum sensing loci used DreamTaq 2X Green PCR Master Mix (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.A.). The PCR protocol applied 50 µL reaction volumes containing final concentrations of: 1X DreamTaq Green PCR Master Mix; 0.3 µM of each primer; ~ 1 ng template DNA; and sufficient nuclease-free water (Thermo Fisher Scientific, Waltham, Massachusetts, U.S.A.) to bring the final volume to 50 µL. The

PCR protocol applied initial denaturation at 95 °C for 2 minutes; and 35 cycles of: denaturation at 95 °C for 30 seconds; annealing at 53 °C for 15 seconds; and extension for 2 minutes at 72 °C; with a final extension of 5 minutes at 72 °C.

The PCR products were separated and visualised using agarose gel electrophoresis. SeaKem LE agarose (Lonza Bioscience, Basel, Switzerland) was used at concentrations of 1 % w/v with 1x Sodium Boric Acid (SBA) buffer (10 mM NaOH and 38 mM Boric acid, adjusted to pH 8.5) (Brody and Kern, 2004). Agarose gels were pre-stained with 1x CondaSafe (Laboratorois Conda, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.). Gel images were captured under ultraviolet light on SynGene G:Box using the Syngene GeneSnap software (version 7.09) and analysed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England).

Amplified gene fragments were sequenced using an ABI 3500XL genetic analyser (Applied Biosystems, California, USA) at Inqaba Biotec Laboratories (Pretoria, South Africa) using the UniComQ forward primer for *B. subtilis* isolates, and the ComBamyR reverse primer for *B. velezensis* isolates. Gene sequences were visualised and edited in Chromas Lite (version 2.01) and consensus sequences generated using BioEdit (version 7.2.6.1) (Hall, 1999). The nucleotide sequences were then submitted to NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Zhang *et al.*, 2000) for comparison to existing gene sequence data in the GenBank database. The BLASTn search parameters for the NCBI nucleotide collection (nr/nt) database were limited to the *B. subtilis* group (Taxid: 653685), omitted uncultured/environmental material, and used the default MEGABLAST algorithm.

4.2.5. Sequencing and phylogeny of *comQ* gene sequence

A *com*Q gene sequence phylogeny was generated using the *comQ* of *B. subtilis* and *B. velezensis* isolates, with counterpart sequences of known pherotype strains (Ansaldi *et al.*, 2002; Tortosa *et al.*, 2001) (Table 4.3.). In this phylogeny, the *comQ* for pherotype NAF4 (*B.*

subtilis subsp. natto) was not available on GenBank, and has been replaced with an identical counterpart from *B. subtilis* subsp. *natto* BEST195 (Nishito *et al.*, 2010).

Table 4.3. Partial quorum-sensing gene loci sequences or whole genomes of *B. subtilis sensu lato* pherotypes selected for this study and sourced from the NCBI GenBank database.

Species and Strain*	GenBank Accession No.
B. subtilis RO-FF-1	AF456130.1
B. subtilis RO-PP-2	AF456131.1
B. subtilis RO-A-4	AF456132.1
B. subtilis RO-DD-2	AF456133.1
B. mojavensis RO-C-2	AF456134.1
B. mojavensis RO-B-2	AF456135.1
B. subtilis RS-D-2	AF456136.1
B. subtilis RO-F-3	AF456137.1
B. subtilis RO-E-2	AF456138.1
B. mojavensis RO-H-1	AY003901.1
B. subtilis RS-B-1	AY003900.1
B. subtilis BEST195 (~NAF4)[#]	NC_017196.2
B. subtilis subsp. *subtilis* 168	NC_000964.3

***** Pherotypes in *Bacillus subtilis sensu lato* as described by Tortosa *et al.*, 2001 and Ansaldi *et al.*, 2002. [#]Applied as counterpart to B. subtilis NAF4 pherotype (Nishito *et al.*, 2010) owing to lack of NAF4 availability on the GenBank database.

Evolutionary relationships between sequences were inferred using MEGA X software (version 10.0.5.) (Kumar *et al.*, 2018) using a sequence alignment performed in MEGA X using MUSCLE (Edgar, 2004). Phylogenies were generated using the Maximum Likelihood method and Kimura 2-parameter model (Kimura, 1980). The initial trees were obtained by Neighbour-Join and BioNJ algorithms which were applied to a pairwise distances matrix, and these were then applied to the heuristic search. This matrix was estimated using the Maximum Composite Likelihood approach and selected the topology with the highest log likelihood value. Phylogeny testing applied a bootstrap method of 500 replicates.

4.2.6. *comQXP* PCR-RFLP pherotyping of *Bacillus subtilis sensu lato* isolates

PCR-RFLP of the *comQXP* gene loci was undertaken for all isolates, as well as reference type strains and pherotype strains. The restriction enzyme choices for this gene sequence were determined with the aid of the online *In Silico* PCR tool using simulated PCR-RFLP (URL: http://insilico.ehu.es/PCR) (San Millán *et al.*, 2013) with the appropriate primer pairs (Chapter 2). PCR amplicons were prepared for restriction digestion using a GeneJet PCR Purification kit (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) as per the manufacturer's instructions and using sterile TE buffer, which comprised 10 mM Tris (pH 8.0) and 1 mM Na_2EDTA (pH 8.0). The final cleaned DNA was quantified and analysed for purity using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) and stored at -20 °C until use.

Four restriction enzymes were applied to this RFLP analysis: *Hpy*166II, *Fnu*4HI, *Bts*CI, and *Cac*8I (New England Biolabs, Ipswich, Massachusetts, U.S.A.) and were utilised as per the manufacturer's recommendations. The digests were carried out in a G-Storm GS1 Thermal Cycler (G-Storm, Somerset, U.K.) as indicated in Table 4.4.. Restriction enzymes *Hpy*166II, *Cac*8I and *Bts*CI were inactivated using a heating step, while *Fnu*4HI required the addition of a stop solution to arrest enzyme digestion. The stop solution was added at 10 µl per 50 µl digestion reaction, and, as per the kit manufacturer's instructions, comprised the following: Ethylenediaminetetraacetic acid (EDTA) 50 mM, bromophenol blue 0.05 % (v/v), and glycerol 50 % (v/v).

Table 4.4. Digestion protocols for restriction enzymes used for *comQXP* PCR-RFLP.

Restriction Enzyme	Digestion	Inactivation
*Fnu*4HI	37 °C for 15 minutes	Stop solution
*Cac*8I	37 °C for 60 minutes	80 °C for 20 minutes
*Bts*CI	50 °C for 15 minutes	80 °C for 20 minutes
*Hpy*166II	37 °C for 15 minutes	65 °C for 20 minutes

The respective restriction digest products were separated and visualised by agarose gel electrophoresis using a 1% (w/v) SeaKem LE Agarose gel (Lonza Bioscience, Switzerland) using 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Band sizes were confirmed using a Fermentas GeneRuler 100 bp Plus DNA Ladder (Fermentas Life Sciences, Waltham, Massachusetts, U.S.A.). Agarose gels were pre-stained with 1x CondaSafe (Laboratorois Conda, Madrid, Spain). Gels were run at 80V for 100 minutes before visualisation. Images of the agarose gel electrophoreses were captured after under ultraviolet light using a SynGene G:Box and Syngene GeneSnap software (version 7.09 and restriction profiles were analysed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Dendrograms were generated using the Syngene GeneTools software using the following parameters: UPGMA (Unweighted Pair Group Method with Arithmetic Mean) with matching parameters based on profile, alignment by molecular weight, and a 0.8% tolerance.

4.2.7. Pherotyping assay of *Bacillus subtilis sensu lato* isolates

Pherotype diversity amongst isolates in relation to known pherotypes was carried out using a modified LacZ reporter gene assay (Miller, 1972), which targeted *srfA* activity as a proxy for ComX binding to the ComP receptor of the tester strain (Tortosa *et al.*, 2001). The assay exploited the LacZ insert in the pherotype tester strains (lacking the ability to synthesise their own ComX) against the ComX-containing conditioned media from the isolates, and generated a baseline activity control by including cognate producer/tester pherotype pairs (Ansaldi *et al.*, 2002) (Table 4.2.). Isolate ComX reactivity to each tester strain was indirectly assessed by the presence of a yellow colouration quantifiable at 420 nm.

4.2.7.1. Conditioned medium preparation

Conditioned medium was obtained from pherotype producer strains (Table 4.2.), *B. subtilis sensu lato* isolates, and reference strains. Cells were grown in starter culture, as mentioned previously (Section 4.2.2.), and appropriate amounts of cell pellets were added into the S7 medium to achieve a final OD_{600} 0.7. Inoculated S7 broth was incubated at 28 °C at 150 rpm in a shaking incubator for 16 h. Conditioned medium was prepared from 10 mL of 16-h-old S7

broth cultures. The broth was centrifuged at 12, 000 *x g* for 5 minutes at 4°C (JA-25.15 rotor, Avanti J-26XPI, Beckman) and the cell pellet discarded. The conditioned medium supernatant was filter-sterilised (0.22 micron) at use.

4.2.7.2. Reporter gene assay pherotyping of isolates

The procedure for this assay followed that of Oslizlo *et al.* (2015) with modifications. The assay was carried out in sterile 96-well flat-bottomed microwell plates (Techno Plastic Products AG, Trasadingen, Switzerland). Tester strains were grown in equal volumes (100 µL) of appropriately-amended S7 broth and conditioned medium. Tester cells were inoculated to achieve a final cell OD_{600} 1.0. Controls were included as follows: Tester-cell only, cognate pherotype producer, S7 broth only. Producers were tested in triplicate. The plates were incubated for 16 h at 28°C in a shaking incubator (200 rpm). Cell growth in the plates was measured spectrophotometrically at 600 nm in a using SpectraMax ABSPlus (Version 2.0.10) (Molecular Devices, San José, California, U. S. A.).

The plates were then centrifuged at 1,800 *x g* for 10 minutes at 4°C with slow deceleration (JS-5.3 rotor, Avanti J-26XPI, Beckman). The supernatant was removed from the wells, and the pellets were resuspended in 200 µL of permeabilisation (Z) buffer. The Z-buffer comprised: $Na_2HPO_4.7H_2O$ 60 mM; $NaH_2PO_4.H_2O$ 40 mM; KCl 100 mM; $MgSO_4$ 1 mM, and adjusted to pH 7.0 with 10 M NaOH. To this buffer, β-mercaptoethanol at 5.6% v/v was added at use. Aliquots of 10 µL of toluene were added to each well and the plates placed on ice for 30 minutes, warmed to 30°C, and the substrate then added (50 µL per well). The substrate for the assay was prepared at use and comprised *ortho*-Nitrophenyl-β-galactoside (ONPG) (Melford Laboratories Ltd, Ipswich, United Kingdom) 10 mM dissolved in a solution of: $Na_2HPO_4.7H_2O$ 60 mM; $NaH_2PO_4.H_2O$ 40 mM, and adjusted to pH 7.0 with 10 M NaOH (Oslizlo *et al.*, 2015). The absorbance of each well was spectrophotometrically measured immediately after substrate addition at 420 nm. Isolate pherotype matches were inferred on the basis of the 420 nm absorbance data and expressed as a percentage of the activity of the respective cognate pherotype pair.

4.3. Results

4.3.1. Sequencing of *comQ* gene locus and phylogeny generation

The *com*Q amplicon sizes after sequencing ranged from 600–950 bp in length. After repeated attempts, amplicons for *B. velezensis* bna81, cce146, and bnn282 failed to return usable *comQ* sequence data and were thus omitted from phylogenetic study. The *comQ* gene sequence phylogeny in Figure 4.1. presents the *B. subtilis* and *B. velezensis* isolates in comparison to known pherotype and reference strain *comQ* sequences. Only four isolates grouped with pherotype counterparts: bnd134 (clade Q-BS1) with *B. subtilis* RO-B-2 , isolates sqo279 and bng215 (Q-BS2) with pherotypes *B. subtilis* RS-D-2 and RO-F-3, and bnd156 (Q-BS3) with *B. subtilis* 168. Isolates bnd136, bng221, sqo271, CT-R67, and B81 isolates were grouped as a single clade M-BS1 with MLSA and as a distinct ecotype (Ecotype 3) (Chapter 3). A distinct subclade is formed amongst these five *B. subtilis* isolates in Q-BS2. The *B. velezensis* isolates clustered into two clades that were distinct from the pherotype reference strains and were considered to be novel pherotypes associated with this species. Furthermore, the *B. velezensis* isolates displayed a greater degree of *comQ* sequence relatedness to previously-described pherotypes than they do to each other. Within the Q-BV1 clade further sub-groupings were evident that could indicate additional pherotype diversity amongst the *B. velezensis* isolates evaluated. The reference strains included in this phylogeny show some unexpected groupings, such as *Bacillus* sp. JS with bnd134 (Q-BS1), and *B. velezensis* model strain FZB42 in a clade distinct from any *B. velezensis* isolates, grouping instead with pherotypes of *B. subtilis* (RO-FF-1 and 168) and *B. mojavensis* (RO-C-2). It is interesting to note that *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM7^{T} forms a unique *comQ* sequence clade, but one which bears some relationship to *B. velezensis* isolates in clade Q-BV2. This data presented as a similarity matrix of *comQ* sequences (Figure 4.2.) demonstrates that there exists some level of sequence similarity which may indicate ComX relatedness or promiscuity amongst the pherotypes towards the *B. subtilis sensu lato* isolates. Isolate bnd156 shows a high sequence similarity towards most pherotypes, reference strains, and *B. subtilis* and *Bacillus* sp. JS isolates, but less to towards the *B. velezensis* isolates. Pherotype 168 demonstrates high sequence similarity to most of the isolates, barring isolate bnd156.

The BLAST search results (Table C1) show the top three sequence matches for the *comQ* amplicons, in which the isolate sequences return similarity percentages of between 92.14–

99.88% for counterpart sequences in the GenBank database. The lowest sequence similarity value (92.14%) is for the sequence of isolate CT-R67, though this comparatively low value may be due to the fact that this was also the shortest sequence obtained from sequencing, and was also of comparatively poor sequence quality. The remaining isolates are all sequences in excess of 700 bp in size, and return values in the 97–99% range of similarity percentage. These similarity values are higher than might be expected for such a hypervariable gene region, but are interesting in that they indicate a substantial level of sequence relatedness to strains in other localities and environments. It is noteworthy that within the *B. subtilis* isolates comprising the Q-BS2 clade, similarity hits to *B. halotolerans* are shown, which is unexpected, though *B. halotolerans* is still a member of the *B. subtilis sensu lato* group under the synonym *B. axarquiensis* (Tindall, 2017). It should be noted that sequence quality for this region was not what may be considered ideal, even after multiple amplification and sequencing attempts.

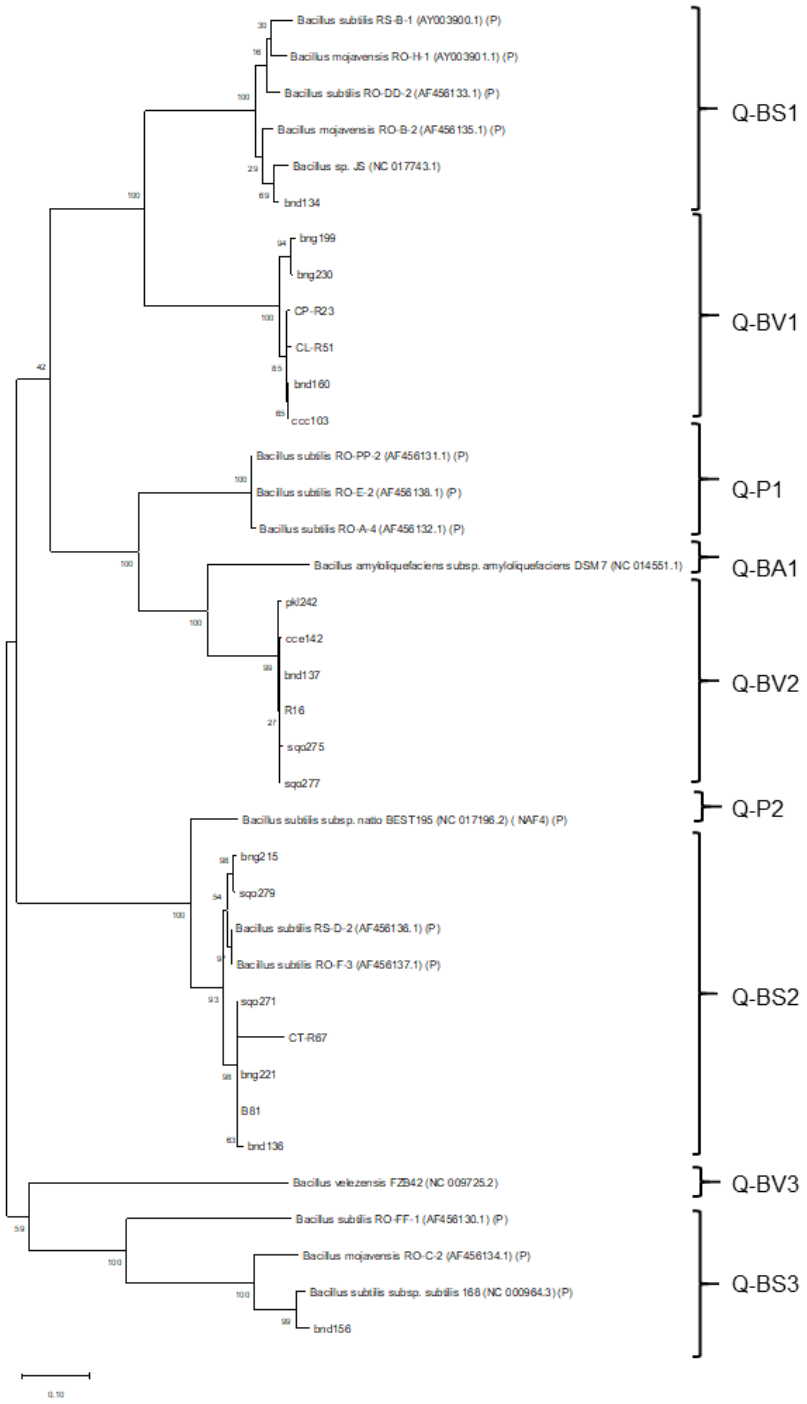


Figure 4.1. Evolutionary analysis of partial *comQ* gene sequences of *Bacillus subtilis sensu lato* isolates and reference strains. The evolutionary analysis was conducted in MEGA X using the Maximum Likelihood method and Kimura 2-parameter model. The tree represents the highest log likelihood (-8048.47). Phylogeny applied 37 nucleotide sequences with 692 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

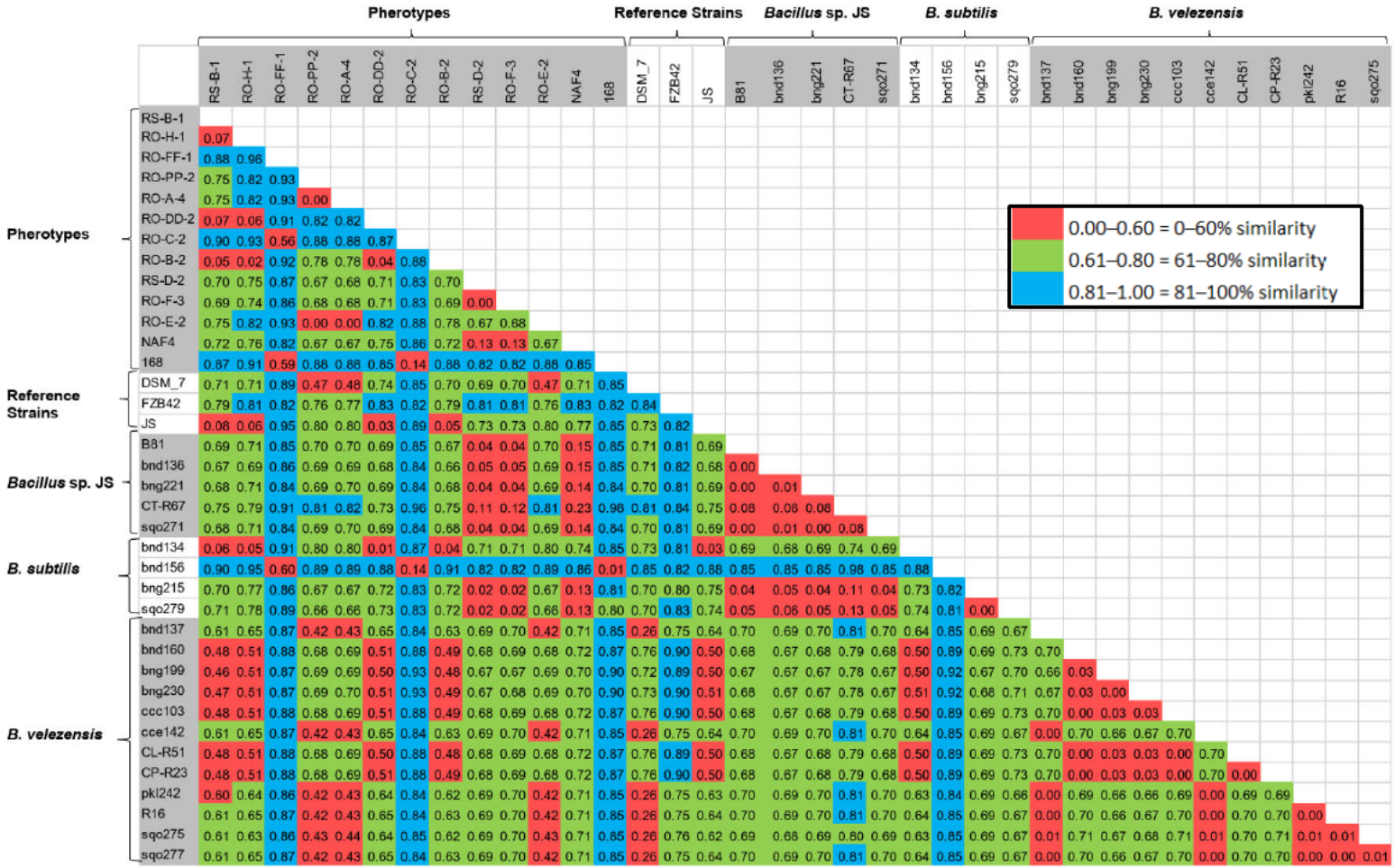


Figure 4.2. Estimates of evolutionary divergence by a pairwise distance matrix of concatenated *comQ* gene sequences for *Bacillus subtilis sensu lato* isolates, reference and pherotype strains. The matrix was generated in MEGA X (Kumar et al., 2018) using the Maximum Composite Likelihood model . The dataset comprised 37 sequences with a total of 509 positions. All positions with less than 95% site coverage were eliminated. Figures indicate the number of base substitutions per site between sequences, heatmap colour grading indicates percentage similarity between sequences as shown in the figure key.

4.3.2. Pherotyping *Bacillus subtilis sensu lato* isolates using *comQXP* PCR-RFLP

An alternative to gene sequencing for the evaluation of putative pherotype variation may be afforded by the PCR-RFLP technique. This approach uses restriction enzymes to digest PCR amplicons into fragments which may be visualised on agarose gel as multiple bands, and patterns may be ascribed to indicate differences between strains. In the context of this study, *comQXP* banding patterns may be interpreted as pherotype variances between isolates and reference strains. The absence of the need for sequence data generation and processing lends this method to larger isolate subsets and may assist in pherotype dereplication when dealing with large numbers of isolates. In this experiment, amplicons of the *comQXP* region for *B. subtilis* and *B. velezensis* isolates and the known pherotype strains were digested using four restriction enzyme candidates: *Bts*CI, *Cac*8I, *Fnu*4HI, and *Hpy*166II. Enzyme *Cac*8I demonstrated incomplete amplicon digestion (Data not shown) and was removed from the enzyme suite.

Enzyme *Bts*CI (Figure 4.3. A and B) demonstrated differing profiles for the *B. subtilis* isolates and pherotypes. The profile of *B. subtilis* subsp. *subtilis* DSM 10^T matched that of a pherotype strain *B. subtilis* subsp. *subtilis* 168, which was expected as these two strains are counterparts. Isolates comprising a *com*Q subclade Q-BS2 (viz., bng221, sqo271, bnd136, CT-R67, and B81) formed a common RFLP profile grouping. In this instance the pherotype fingerprinting demonstrated a good correlation with *comQ* sequence data, and with MLSA and ES (Chapter 3). The remaining *B. subtilis* isolates did not present any significant similarities to each other or any of the pherotype reference strains, which closely mirrors the findings of the *comQ* phylogeny. For the *B. velezensis* isolates, *Bts*CI allowed grouping of the isolates across two major profile groups, with the only divergent profile belonging to cce146. This was correlated with the ES finding that cce146 was a unique ecotype amongst the isolate subset. The two major profile groupings reflected the isolate groupings observed in *com*Q clades Q-BV1 and Q-BV2, respectively (Figure 4.1.). None of the *B. velezensis* profiles match those of the pherotype reference strains. The gel images for *Bts*CI are shown in Appendix C (Figures C1 A and B).

Enzyme *Fnu*4HI (Figure 4.4. A and B) distinguished a major grouping amongst the *B. subtilis* isolates, also including pherotype reference RO-FF-1. This major grouping excluded only isolates bnd134 and bnd156. Here again the *B. subtilis* subsp. *subtilis* DSM 10^T profile matched that of pherotype strain *B. subtilis* subsp. *subtilis* 168. The remaining pherotype reference strains demonstrated no significant profile relationship to the isolates or each other. The *B. velezensis* isolates were grouped across three profiles, which again corroborated the *comQ* phylogeny clades, but was further able to distinguish bng199 and bng230 as having differing banding profiles from the remainder of the BV1 clade isolates. The gel images for *Fnu*4HI are shown in Appendix C (Figures C2. A and B).

Enzyme *Hpy*166II displayed incomplete digestion for *B. subtilis* amplicons (data not shown). The RFLP profiles for *B. velezensis* isolates using *Hpy*166II (Figure 4.5.) displayed remarkable similarity to the profiles generated for *Bts*CI, and also grouped the isolate profiles in a manner corroborating the Q-BV1 and Q-BV2 *com*Q clades. This was unexpected, as in silico PCR-RFLP simulations predicted different profiles between these enzymes (Chapter 2). The gel image for *B. velezensis* profiles for enzyme *Hpy*166II are shown in Appendix C (Figure C3).

In terms of the potential for PCR-RFLP as a rapid-pherotyping technique, restriction enzyme *Bts*CI may be considered to be better suited to defining putative pherotype groupings amongst the *B. subtilis* isolates, as it grouped profiles from those isolates comprising clade Q-BS2. Both *Bts*CI and *Fnu*4HI demonstrated the ability to define putative pherotypes amongst *B. velezensis*, although *Fnu*4HI demonstrates increased sensitivity towards very closely-related *comQ* sequences. Reference strain *B. subtilis* subsp. *subtilis* DSM 10^T only representative *B. subtilis sensu lato* which could be applied to PCR-RFLP validation of the enzymes identified for this purpose in silico (Chapter 2). The PCR-RFLP fingerprints for this reference strain in *Bts*CI and *Fnu*4HI closely match those for the *B. subtilis* strain represented in the in silico tool, indicating that the in silico approach was useful for evaluating enzyme candidates.

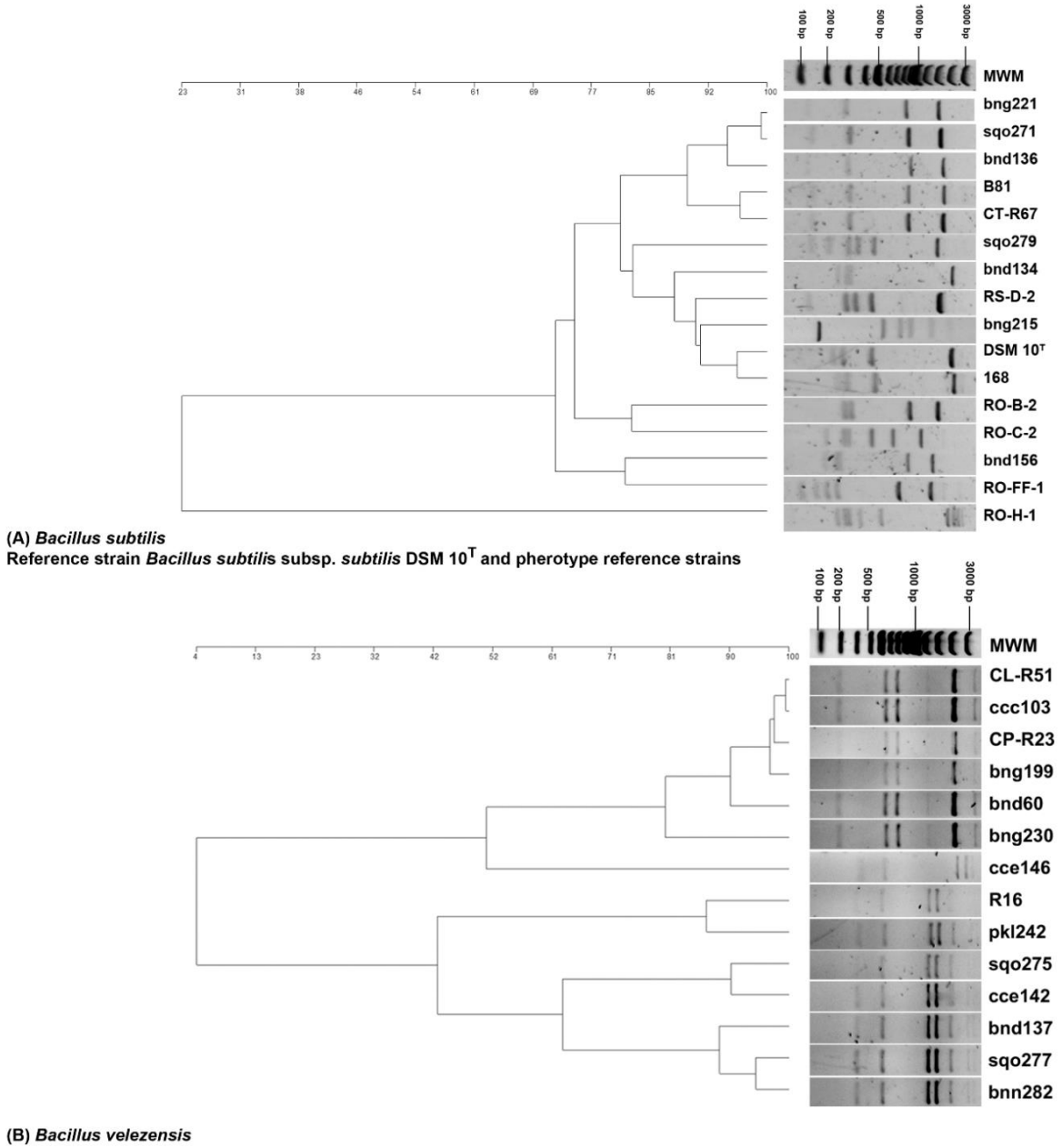


Figure 4.3. Dendrograms evaluating relationships between restriction digest profiles generated from *comQXP* PCR-RFLP analysis using the *Bts*CI restriction enzyme. Profiles shown include the plant-associated *B. subtilis sensu lato* isolates and reference type strain *B. subtilis* DSM 10^T, and pherotype reference strains. Fingerprint analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint dendrogram parameters as follows: UPGMA with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with a 0.8% tolerance. "MWM" denotes the 100 bp Plus DNA ladder (Fermentas Life Sciences, Waltham, Massachusetts, U.S.A.) included in the agarose gels. Gel images have been inverted for clarity.

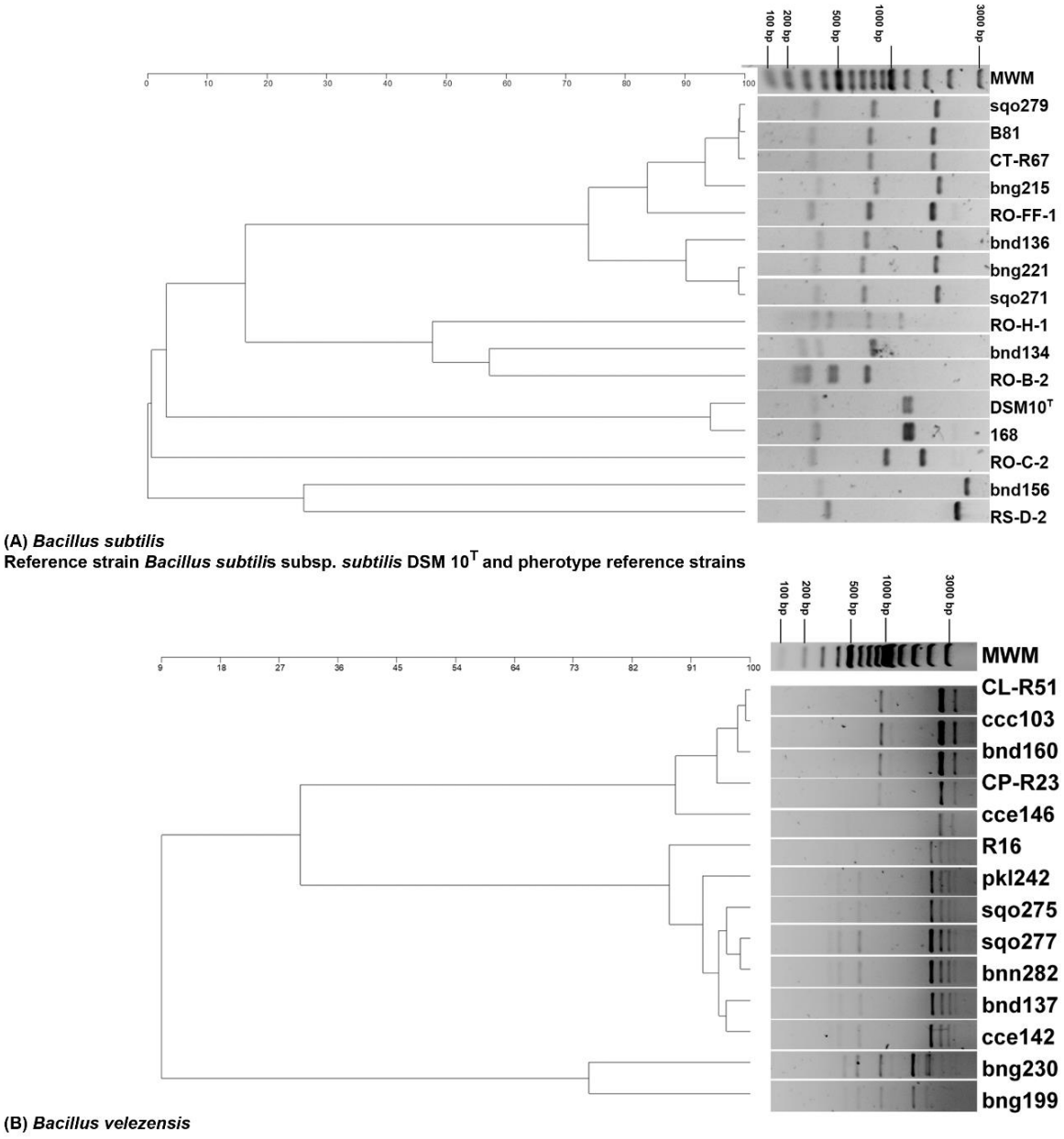


Figure 4.4. Dendrograms evaluating relationships between restriction digest profiles generated from *comQXP* PCR-RFLP analysis using the *Fnu*4HI restriction enzyme. Profiles shown include the plant-associated *B. subtilis sensu lato* isolates and reference type strain *B. subtilis* DSM 10^T, and pherotype reference strains. Fingerprint analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint dendrogram parameters as follows: UPGMA with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with a 0.8% tolerance. "MWM" denotes the 100 bp Plus DNA ladder (Fermentas Life Sciences, Waltham, Massachusetts, U.S.A.) included in the agarose gels. Gel images have been inverted for clarity.

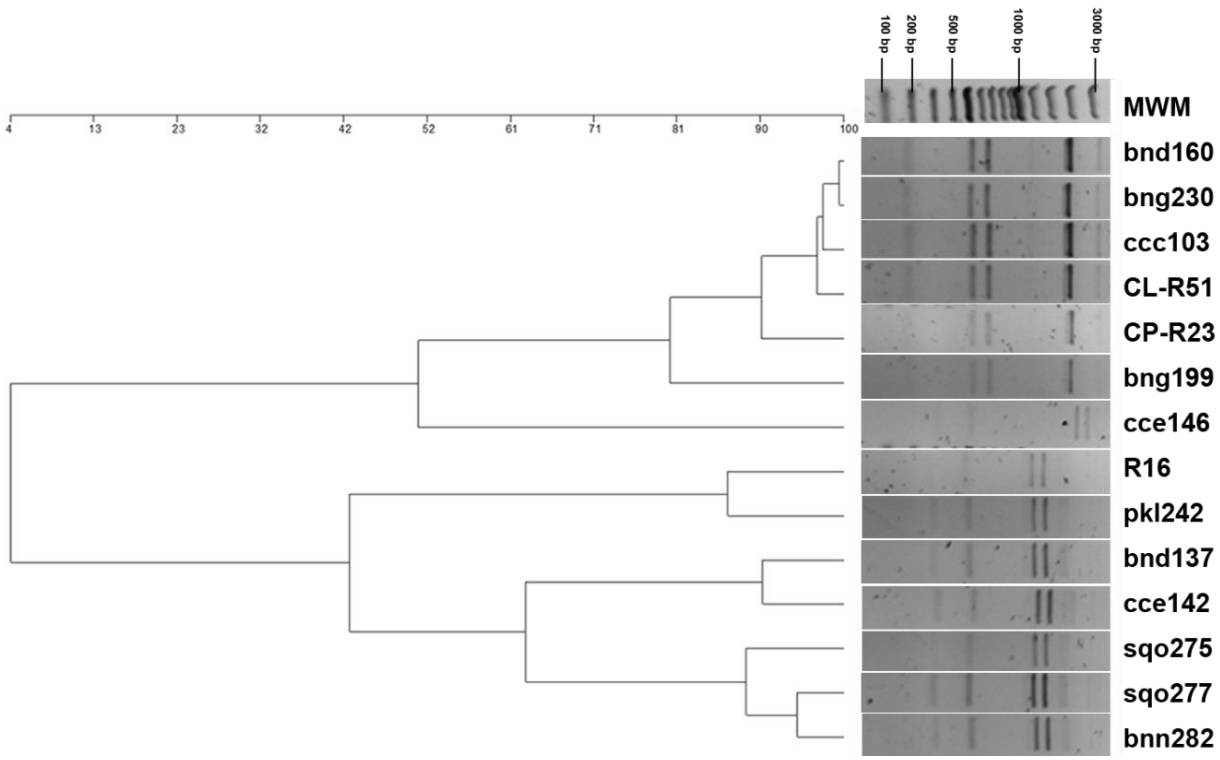


Figure 4.5. Dendrograms evaluating relationships between restriction digest profiles of *Bacillus velezensis* isolates generated from *comQXP* PCR-RFLP analysis using the *Hpy*166II restriction enzyme. Profiles shown include the plant-associated *B. velezensis* isolates. Fingerprint analyses were performed using Syngene GeneTools software (version 4.3.8.0.) (Syngene, Cambridge, England). Fingerprint dendrogram parameters as follows: UPGMA with matching parameters using profile, Jaccard similarity coefficient, and alignment by molecular weight with a 0.8% tolerance. "MWM" denotes the 100 bp Plus DNA ladder (Fermentas Life Sciences, Waltham, Massachusetts, U.S.A.) included in the agarose gels. Gel images have been inverted for clarity.

4.3.3. Isolate pherotyping using *srfA*-LacZ reporter gene assay

Pherotyping assays were carried out using ComX-containing conditioned medium to stimulate pherotype tester strains carrying a *srfA*-LacZ reporter gene, allowing tester strain stimulation to be interpreted as a proxy for ComX reactivity and to be quantified as an absorbance value at 420 nm after addition of ONPG substrate. In this assay, the 420 nm absorbance value was recorded and expressed as percentage of that achieved by the cognate producer/tester pherotype pair. As such, a percentage value of 100% (or greater, as noted in some cases) was interpreted to constitute a pherotype match. The assays had been optimised (Data not shown) to apply sufficient tester cell density to ensure that the cognate pherotype producer/tester pairs returned values A_{420} ~0.4 after substrate addition. As a consequence of tester cell density, many of the percentage activity ranges for the isolates were high, and demonstrated stimulation of the tester strain that exceeded the level achieved by the cognate producer/tester pherotype pair.

In several cases, stimulation of the tester strain at levels exceeded 100%, and these isolates indicated matches to multiple pherotypes. To this end, a rating scale was introduced to aid in defining pherotype matches, with percentage activity ratings ascribed as follows: (-) < 20% (negative response), (±) 21–50%, (+) 51–80%, (++) 81–99%, and (+++) indicating 99–100% or greater (i.e. pherotype match). These summarised results of the isolates against the tester strains are presented in Table 4.5. For clarity, graphical representations of the isolates' percentage activity against the seven pherotype strains are presented in Figure 4.6.

Amongst *B. subtilis* isolates (Table 4.5. A and Figure 4.6. A) B81, bnd156, and sqo279 matched to multiple pherotypes. Reference strain *B. subtilis* subsp. *subtilis* DSM 10^T did not match to the expected counterpart pherotype 168, and in fact failed to match significantly to any tester strains. On the basis of the rating system, certain *B. subtilis* isolates did not demonstrate any significant level of tester strain stimulation, and these were interpreted to have no pherotype match: bnd136, bng221, and sqo271. The assay data demonstrated poor correlation to the *comQ* phylogeny groupings distinguished. Isolates which did match to pherotypes (viz., bnd134 and CT-R67) did not match with those pherotypes that were expected based on gene sequence data.

The *B. velezensis* isolates and reference strains achieved overall higher levels of tester strain response than did the *B. subtilis* isolates (Table 4.5. B and Figure 4.6. B) . Four *B. velezensis* isolates did not match to any tester strain pherotypes: bnd137, bnd136, ccc103, and CP-R23. All of the remaining *B. velezensis* isolates matched to multiple pherotypes, with matches to three tester strains being widely observed: RO-C-2, RO-H-1, and 168. The reference strains *B. velezensis* DSM 23117 ^{T} and *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T} each matched to a range of pherotypes, with DSM 7^{T} matching to all pherotypes but RO-FF-1.

Table 4.5. Pherotype tester strains responses to *Bacillus subtilis* and *B. velezensis* isolate conditioned media as determined by a *srfA*-LacZ reporter gene assay.

(A) *Bacillus subtilis* isolates and reference type strain *B. subtilis subsp. subtilis* DSM 10^T.

Strain/Isolate	Pherotype Tester Activity*							Pherotype Matches#
	RO-B-2	RO-E-2	RO-C-2	RS-D-2	RO-H-1	RO-FF-1	168	
DSM 10^T	+	++	++	+	++	+	++	None
B81	±	+++	+++	+	++	±	+++	RO-E-2, RO-C-2, 168
bnd134	+	+	++	++	++	±	+++	168
bnd136	+	+	+	+	+	±	++	None
bnd156	±	+	+++	+	+++	±	++	RO-C-2, RO-H-1
bng215	+	+	+	+	+++	±	++	RO-H-1
bng221	±	+	+	+	+	±	+	None
CT-R67	+	+	+	+	+	+	+++	168
sqo271	+	+	+	±	+	±	+	None
sqo279	±	+++	+++	+++	++	±	+	RO-E-2, RS-D-2, RO-C-2

*Activity rating scale represents percentage activity ratings ascribed as follows: (-) < 20% (negative response), (±) 21–50%, (+) 51–80%, (++) 81–99%, and (+++) indicating 99–100% or greater. # Rating (+++) interpreted as a pherotype match.

Table 4.5. Continued.

(B) *Bacillus velezensis* isolates and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T} and *B. velezensis* DSM 23117^{T}.

Strain/Isolate	Pherotype Tester Activity*							Pherotype Matches#
	RO-B-2	RO-E-2	RO-C-2	RS-D-2	RO-H-1	RO-FF-1	168	
DSM 23117 ^{T}	+	+++	+++	+++	+++	+	+	RO-E-2, RS-D-2, RO-C-2, RO-H-1
DSM 7 ^{T}	+++	+++	+++	+++	+++	+	+++	RO-B-2, RO-E-2, RO-C-2, RS-D-2, RO-H-1, 168
bna81	±	±	++	+	+++	-	+++	RO-H-1, 168
bnd137	±	+	++	+	++	±	++	None
bng160	-	±	-	-	+	-	±	None
bng199	+	++	+++	+	+++	±	+++	RO-C-2, RO-H-1, 168
bng230	+	+++	+++	++	+++	±	+++	RO-E-2, RO-C-2, RO-H-1, 168
bnn282	±	++	++	+	+++	±	+++	RO-H-1, 168
ccc103	±	±	-	±	±	-	-	None
cce142	+	++	+++	++	+++	±	++	RO-C-2, RO-H-1, 168
cce146	+	+	+	+	++	+	+	None
CL-R51	+	+	+++	+	+++	±	+++	RO-C-2, RO-H-1, 168
CP-R23	-	-	-	-	-	-	-	None
pkl242	+	++	+++	++	+++	±	+++	RO-C-2, RO-H-1, 168
R16	±	+	++	+	+++	±	+++	RO-H-1, 168
sqo275	+	++	+++	++	+++	±	+++	RO-C-2, RO-H-1, 168
sqo277	±	+	+++	+	+++	+	+++	RO-C-2, RO-H-1, 168

*Activity rating scale represents percentage activity ratings ascribed as follows: (-) < 20% (negative response), (±) 21–50%, (+) 51–80%, (++) 81–99%, and (+++) indicating 99–100% or greater. # Rating (+++) interpreted as a pherotype match.

Figure 4.6. Graphical representation of isolate ComX activity against known pherotype tester strains. Activation of the ComQXPA signalling cassette for tester strains of seven pherotypes was assessed using a *srfA*-LacZ reporter gene assay. The response of tester strains after incubation with conditioned medium from *B. subtilis* and *B. velezensis* isolates and reference strains was expressed as a percentage of the activity observed for the cognate pherotype tester/producer pair (viz., RO-B-2, RO-E-2, RO-C-2, RS-D-2, RO-H-1, RO-FF-1, and 168).

(A) *Bacillus subtilis* isolates and reference type strain *B. subtilis subsp. subtilis* DSM 10^{T}.

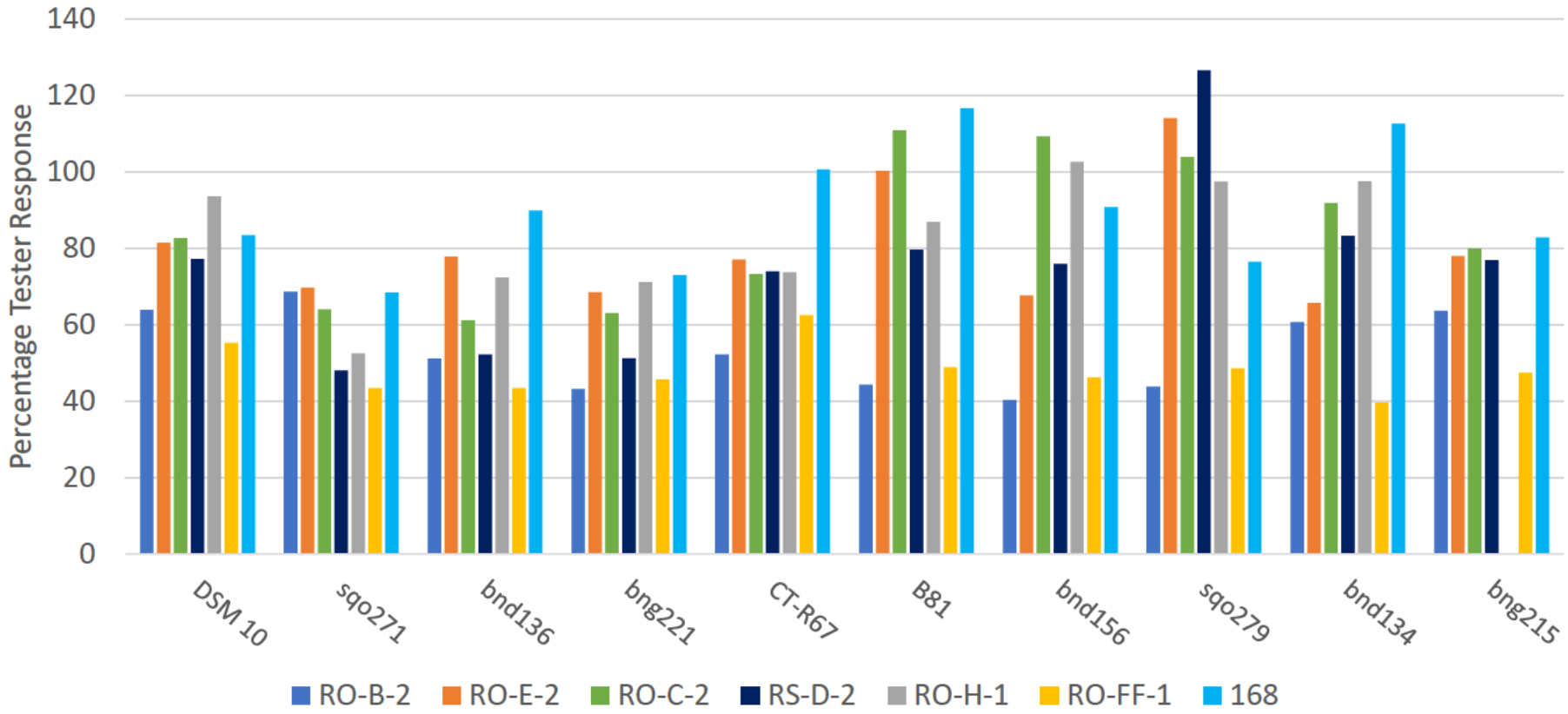
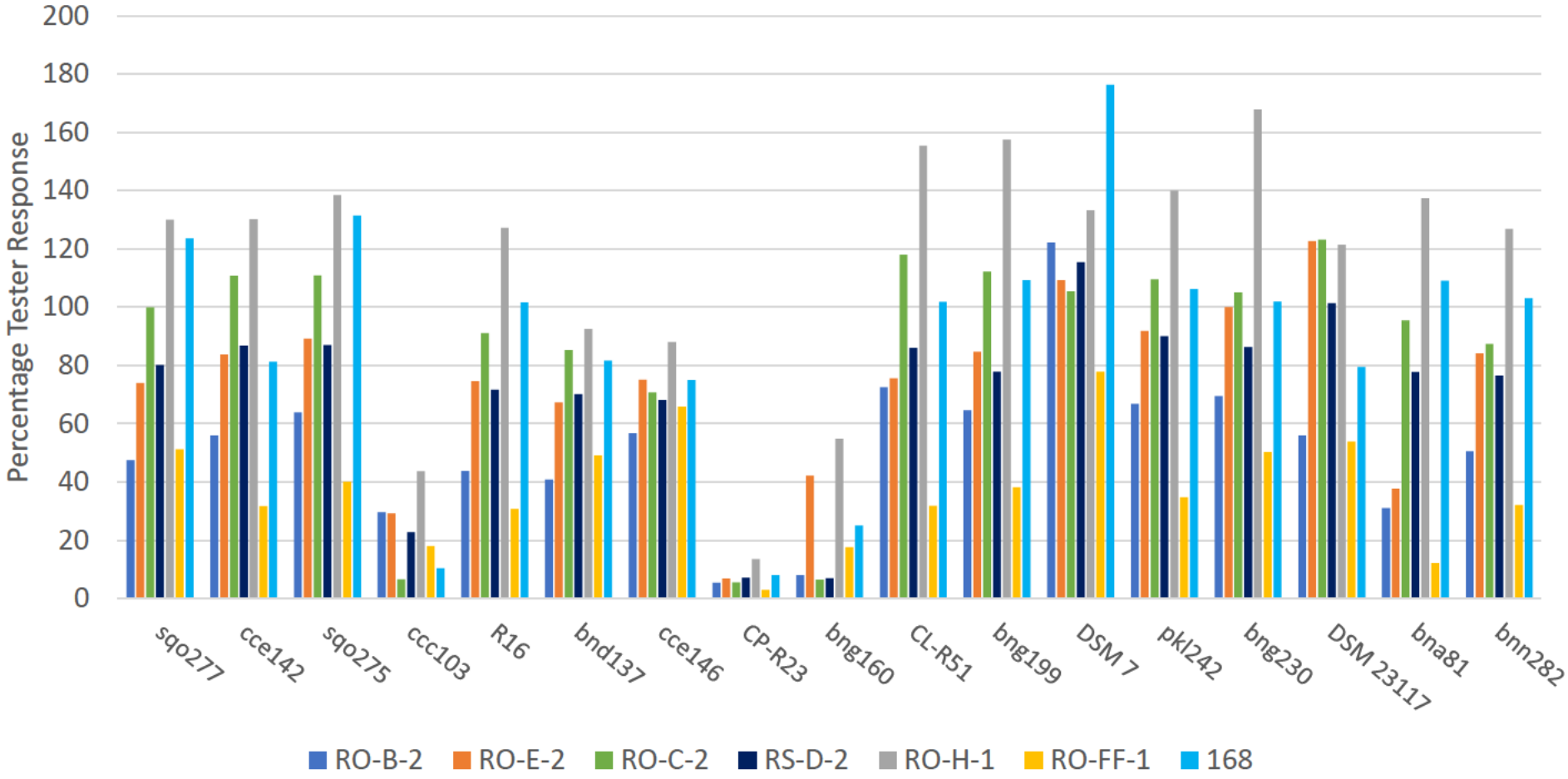


Figure 4.6. Continued.

(B) *Bacillus velezensis* isolates and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T} and *B. velezensis* DSM 23117^{T}.



4.4. Discussion

The *com*Q gene sequence is a region of the *comQXPA* quorum-sensing cassette that has been shown to reflect ComX variation, as this region encodes the isoprenyl transferase responsible for the post-translational modification in ComX (Tortosa *et al.*, 2001). The *comQ* gene sequence phylogeny highlighted that the *B. velezensis* isolates' *comQ* sequences were distinct from those clusters *of B. subtilis* isolates and pherotype reference strains, and therefore demonstrated concurrent variability in ComX. The sequence clusters also suggested that there were two major pherotypes at play amongst the *B. velezensis* isolates, as further evidenced these groups' restriction digest profiles. In comparison to *B. velezensis* DSM 23117^{T}, a commercialised European-derived plant-associated strain FZB42, the South African isolates demonstrated some level of pherotype predominance not related to this plant-associated *B. velezensis* model strain (Fan *et al.*, 2018). Reference strain DSM 23117^{T} also demonstrates some relatedness to *B. subtilis* and *B. mojavensis* pherotypes, which is suggestive of some degree of pherotype interspecies relatedness. The study isolates originated from different crop types and locations, but show some degree of similarity in pherotype match ranges and quorum-sensing gene sequences. The isolates comprising *comQ* clades Q-BV1, Q-BV2, and Q-BS2 originated from different crop types and localities, and that the isolates in these respective groups demonstrated such a degree of similarity in pherotype gene loci is noteworthy. Those isolates comprising an ecotype with *Bacillus* sp. JS (Chapter 3) formed a cluster (Q-BS2) that is distinct to this reference strain the *comQ* phylogeny. Phylogenetic investigations of *comQ* or *comQXP* may be applied as indirect indicators of ComX variation, but do not indicate whether the pheromone variants endow any cross-species communication or are promiscuous between pherotype groups. The *comQXP* PCR-RFLP carried out using four restriction enzyme candidates determined enzymes *Bts*CI and *Fnu*4HI to be useful for rapid pherotyping, as these enzymes were able to group isolates in a manner reflective of *comQ* clustering. A PCR-RFLP *comQXP* has been reported previously by Tortosa *et al.* (2001), although this method was applied to confirm that pherotype producer strains contained the target modifications, rather than as a means of determining pherotypes within a subset of unknown isolates, which was the intention of the method as applied here.

The sequence-based studies of putative pherotype variation amongst isolates yielded a clearer picture of the pherotypes prevalent amongst the isolate subset. The pherotyping assay failed to corroborate the isolate pherotypes as expected based on the *comQ* sequence-based clusters. For example, isolate bnd156 formed a clade with pherotype *B. subtilis* 168 in the *comQ* phylogeny, but matched with pherotypes RO-H-1 and RO-C-2 in the assay. Permutations of the microplate pherotyping assay as applied in this study have been reported elsewhere and these studies had reported corroboration of *comQ* data and pherotyping assay (Oslizlo *et al.*, 2015; Stefanic *et al.*, 2012; Stefanic and Mandić-Mulec, 2009). It was surprising to note in this study that several isolates matched to multiple pherotypes in the assay. The assay methodology may be inherently limited, as it would not account for any cross-reactivity capability arising from isolates' ComX post-translational modifications (particular *B. velezensis*) that may allow a greater degree of ComX variant cross-reactivity. Furthermore, ComX concentrations in the conditioned medium produced by the isolates cannot be quantified, and this may cause excessive stimulation of tester strains. Some level of cross-reactivity may be expected amongst the pherotype strains, based on previous studies (Mandić-Mulec *et al*., 2003), and this may be attributed to the nature of the pheromone isoprenylation and side chain (Okada, 2011). In the pherotyping assay the *B. velezensis* isolates demonstrated promiscuity amongst the pherotypes, in particular RO-C-2, RO-H-1, and 168. Pherotypes RO-C-2 and 168 have demonstrated cross-reactivity in previous studies, but RO-H-1 has not been shown to react to unrelated ComX variants (Mandić-Mulec *et al*., 2003). Promiscuity and cross-species communication have been reported in ComX-mediated quorum sensing (Mandić-Mulec *et al*., 2003, Ansaldi *et al*., 2002), although a higher resolution assay and investigation of the make-up of isolates' ComX variants may shed some light as to the extent of these occurrences within this subset of plant-associated isolates.

Comparisons of *comQ* sequence similarity and pherotype assay performance highlight some differences between the two methods in terms of the applicability of gene sequence to predict matching assay outcomes. Where there was high *comQ* gene sequence similarity, it could be anticipated that higher levels of ComX activity would be observed for that pairing in the pherotype assay. This does not appear to be the case in this study, for example: RO-FF-1 sequence similarity suggests a high level of promiscuity with this tester strain, which was not observed in the assay data. Likewise, RO-H-1 demonstrated high promiscuity in the assay, but showed varying levels of *comQ* sequence similarity to the isolates. The incongruence between these two pherotyping approaches (viz. *comQ* gene sequence and pherotype assay) may be due

to the pherotyping assay requiring further optimisation to allow better resolution of pherotype matching. Furthermore, based on *comQ* gene sequence similarity inferences, a study of the activity levels of ComX amongst the plant-associated isolates would be of interest.

The significant differences between pherotype responses and the expected interactions based on gene sequence data suggest that some degree of assay optimisation may be required in order to fine-tune the assay to suit environmental isolates, and in particular for application in *B. velezensis* pherotyping. Aspects such as growth medium and nutrient amendments may also need to be addressed to better suit isolate requirements and any engineered auxotrophies. The assay had been previously optimised to use a tester cell density that suited the cognate pherotype producer/tester pairs (Data not shown), and this was clearly too high a cell density to allow resolution of environmental isolate pherotypes. The cognate pherotype producer strains—perhaps as a consequence of their genetic manipulation and/or domestication—seemed to produce lower levels of ComX than the environmental strains. To this end, greater amounts of tester cells needed to be applied to gain achieve ONPG absorbance values that could be considered trustworthy ($A_{420} \geq 0.4$). This would be difficult to account for in future assay optimisation, as these cognate pairs form the baseline against which responses of tester to environmental isolates are measured, and clearly either the tester or producer (i.e. isolate) cell concentrations applied in this study were too high and resulted in excessive tester strain responses. Another avenue would be to extract the isolate subset ComX peptides and perform the assay using standardised amounts of purified pheromone (Bacon Schneider *et al*., 2002) rather than to apply culture supernatant with potentially variable concentrations of ComX. Purified isolate pheromones would further be amenable to chemical characterisation of the ComX variants' structures and compositions (Ansaldi *et al*., 2002), which would afford confirmation of the gene-sequence-based differences observed amongst the isolates, and perhaps explain the pheromone promiscuity observed in the pherotype assay.

The pherotype assay findings may be interpreted to represent a high degree of species-level ComX cross-reactivity in the interests of common goods (e.g. Surfactin) in the plant-associated environment, as the selection criteria for these isolates from environmental samples was biased towards surfactin production (Tredgold, 2015, Hunter 2016). A possibility may be put forward

that *B. velezensis* ComX is less discrete and more promiscuous than those of the *B. subtilis* isolates. If this is the case, then a question remains as to the exact ecological role ComX cross-reactivity could be playing in the plant-associated environment. Perhaps such ComX cross-reactivity could represent a form of species-level communication in addition to CSF (Pottathil *et al.*, 2008) which allows interactions between a greater range of pherotypes than may be expected in previously-examined environments (Oslizlo *et al.*, 2015; Stefanic and Mandić-Mulec, 2009). Alternatively, these promiscuous pherotypes may be the dominant colonists on the plant host—as evidenced by their presence across a range of crop types and sampling locations—and this could serve to remove competition from other less-beneficial or less-competitive *Bacillus* species.

This study presented the first report of *B. velezensis* pherotyping, and also demonstrated that these plant-associated *B. velezensis* showed divergence in pherotype from previously-described *B. subtilis sensu lato* pherotypes. Furthermore, the overall levels of tester strain stimulation achieved by *B. velezensis* isolates were higher than those observed for the *B. subtilis* isolates, and these isolates may demonstrate some degree of pherotype cross-reactivity. Ultimately, studies of the isolates' ComX isoprenylation and side chain composition is warranted in order to gain a full picture of pherotype variance. Future avenues for research with these isolates may also examine kin discrimination amongst the subset, as this has been reported to vary between ecotypes and impact co-habitation on plant host surfaces (Stefanic *et al.*, 2015).

4.4.1. Conclusions

This study presented three methods by which ComX pherotypes may be defined in plant-associated isolates of *B. subtilis sensu lato*: Indirectly using sequence-based variation in *comQ* and the *comQXP* PCR-RFLP, and directly using ComX reactivity in a *srfA*-LacZ reporter gene assay. Similarities could be drawn in the findings of the *com*Q gene sequence phylogeny and *comQXP* PCR-RFLP. In the *com*Q phylogeny, three clusters of isolates did not group with known pherotypes, and these were interpreted to constitute novel ComX variants. Two such clusters were observed in the *B. velezensis* isolates, and one amongst the *B. subtilis*. Only four isolates demonstrated significant relationships to known pherotypes, and these belonged to *B. subtilis*. Of the four restriction enzyme candidates evaluated for the *comQXP* PCR-RFLP,

*Bts*CI and *Fnu*4HI were determined to be the most compatible for the *comQXP* region of plant-associated *B. subtilis sensu lato*. Enzyme *Bts*CI demonstrated greater sensitivity to *B. subtilis*, displaying profile groups that were most similar to the *comQ* phylogeny, in particular the clade that comprised isolates B81, sqo271, bng221, and bnd136 and CT-R67. Enzyme *Fnu*4HI afforded greater resolution of closely-related *B. velezensis* isolate pherotypes on the basis of comparison to *comQ* phylogenetic clades. The reporter gene assay yielded results that were not consistent with either sequence-based method, but demonstrated that the *B. velezensis* isolates showed very high levels of activity compared to the cognate producer/tester pherotype interactions, and also suggested that these isolates demonstrate levels of pherotype promiscuity. Overall, the findings from this pherotyping study suggested that the plant-associated *B. subtilis* and *B. velezensis* pherotypes show degrees of localised distinctness to known *B. subtilis sensu lato* pherotypes, particularly amongst the *B. velezensis* isolates.

4.5. References

Ansaldi, M. & Dubnau, D. (2004) Diversifying selection at the *Bacillus* quorum-sensing locus and determinants of modification specificity during synthesis of the ComX pheromone. *Journal of Bacteriology* **186**: 1, 15–21. 10.1128/jb.186.1.15-21.2004

Ansaldi, M., Marolt, D., Stebe, T., Mandić-Mulec, I. & Dubnau, D. (2002) Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Molecular Microbiology* **44**: 6, 1561–1573. https://doi.org/10.1046/j.1365-2958.2002.02977.x

Bacon Schneider, K., Palmer, T. M. & Grossman, A. D. (2002) Characterization of *comQ* and *comX*, two genes required for production of ComX pheromone in *Bacillus subtilis*. *Journal of Bacteriology* **184**: 2, 410–419. 10.1128/JB.184.2.410-419.2002

Brody, J. R. & Kern, S. E. (2004) Sodium boric acid: A Tris-free, cooler conductive medium for DNA electrophoresis. *BioTechniques* **36**: 2, 214–216. 10.2144/04362BM02

Camilli, A. & Bassler, B. L. (2006) Bacterial small-molecule signaling pathways. *Science* **311**: 5764, 1113–1116. 10.1126/science.1121357

Connor, N., Sikorski, J., Rooney, A. P., Kopac, S., Koeppel, A. F., Burger, A., Cole, S. G., Perry, E. B., Krizanc, D., Field, N. C., Slaton, M. & Cohan, F. M. (2010) Ecology of speciation in the genus *Bacillus*. *Applied and Environmental Microbiology* **76**: 5, 1349–1358. 10.1128/aem.01988-09

Edgar, R. C. (2004) MUSCLE: Multiple alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 5, 1792–1797. 10.1093/nar/gkh340

Fan, B., Wang, C., Song, X., Ding, X., Wu, L., Wu, H., Gao, X. & Borriss, R. (2018) *Bacillus velezensis* FZB42 in 2018: The Gram-positive model strain for plant growth promotion and biocontrol. *Frontiers in Microbiology* **9**: 2491. 10.3389/fmicb.2018.02491

Hall, T. A. (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98. 10.14601/Phytopathol_Mediterr-14998u1.29

Hamoen, L. W., Venema, G. & Kuipers, O. P. (2003) Controlling competence in *Bacillus subtilis*: Shared use of regulators. *Microbiology* **149**: 9–17. 10.1099/mic.0.26003-0

Hunter, C. H. (2016) Screening of aerobic endospore-forming bacterial isolates as candidate biocontrol agents against *Rhizoctonia solani*. Ph.D. Thesis Thesis, University of KwaZulu-Natal, Pietermaritzburg, KwaZulu-Natal, South Africa.

Joint, I., Allan Downie, J. & Williams, P. (2007) Bacterial conversations: Talking, listening and eavesdropping. An introduction. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **362**: 1483, 1115–1117. 10.1098/rstb.2007.2038

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 2, 111–120. 10.1007/bf01731581

Koeppel, A. F., Wertheim, J. O., Barone, L., Gentile, N., Krizanc, D. & Cohan, F. M. (2013) Speedy speciation in a bacterial microcosm: New species can arise as frequently as adaptations within a species. *The ISME Journal: Multidisciplinary Journal of Microbial Ecology* **7**: 6, 1080–1091. 10.1038/ismej.2013.3

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 6, 1547–1549. 10.1093/molbev/msy096

Mandić-Mulec, I., Kraigher, B., Cepon, U. & Mahne, I. (2003) Variability of the quorum sensing system in natural isolates of *Bacillus* sp. *Food Technology and Biotechnology* **47**: 1, 23–28.

Miller, J. H. (1972) *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, [Cold Spring Harbor, N.Y.

Nishito, Y., Osana, Y., Hachiya, T., Popendorf, K., Toyoda, A., Fujiyama, A., Itaya, M. & Sakakibara, Y. (2010) Whole genome assembly of a natto production strain Bacillus subtilis natto from very short read data. *BMC Genomics* **11**: 243–243. 10.1186/1471-2164-11-243

Okada, M. (2011) Post-translational isoprenylation of tryptophan. *Bioscience, Biotechnology, and Biochemistry* **75**: 8, 1413–1417. 10.1271/bbb.110087

Oslizlo, A., Stefanic, P., Vatovec, S., Beigot Glaser, S., Rupnik, M. & Mandić-Mulec, I. (2015) Exploring ComQXPA quorum-sensing diversity and biocontrol potential of *Bacillus* spp. isolates from tomato rhizoplane. *Microbial Biotechnology* **8**: 3, 527–540. 10.1111/1751-7915.12258

Pottathil, M., Jung, A. & Lazazzera, B. A. (2008) CSF, a species-specific extracellular signaling peptide for communication among strains of *Bacillus subtilis* and *Bacillus mojavensis*. *Journal of Bacteriology* **190**: 11, 4095–4099. 10.1128/jb.00187-08

Reva, O. N., Dixelius, C., Meijer, J. & Priest, F. G. (2004) Taxonomic characterization and plant colonizing abilities of some bacteria related to *Bacillus amyloliquefaciens* and *Bacillus subtilis*. *FEMS Microbiology Ecology* **48**: 249–259. 10.1016/j.femsec.2004.02.003

San Millán, R. M., Martínez-Ballesteros, I., Rementeria, A., Garaizar, J. & Bikandi, J. (2013) Online exercise for the design and simulation of PCR and PCR-RFLP experiments. *BMC Research Notes* **6**: 1, 513. 10.1186/1756-0500-6-513

Sikorski, J. & Nevo, E. (2005) Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 44, 15924–15929. 10.1073/pnas.0507944102

Solomon, J. M., Lazazzera, B. A. & Grossman, A. D. (1996) Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes & Development* **10**: 16, 2014–2024. 10.1101/gad.10.16.2014

Solomon, J. M., Magnuson, R., Srivastava, A. & Grossman, A. D. (1995) Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes & Development* **9**: 5, 547–558. 10.1101/gad.9.5.547

Stefanic, P. & Mandić-Mulec, I. (2009) Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. *Journal of Bacteriology* **191**: 6, 1756–1764. 10.1128/jb.01290-08

Stefanic, P., Kraigher, B., Lyons, N. A., Kolter, R. & Mandic-Mulec, I. (2015) Kin discrimination between sympatric *Bacillus subtilis* isolates. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 45, 14042–14047. 10.1073/pnas.1512671112

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F. M. & Mandić-Mulec, I. (2012) The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. *Environmental Microbiology* **14**: 6, 1378–1389. 10.1111/j.1462-2920.2012.02717.x

Tindall, B. J. (2017) The consequences of *Bacillus axarquiensis* Ruiz-García *et al*. 2005, *Bacillus malacitensis* Ruiz-García *et al.* 2005 and *Brevibacterium halotolerans* Delaporte and Sasson 1967 (Approved Lists 1980) being treated as heterotypic synonyms. *International Journal of Systematic and Evolutionary Microbiology* **67**: 1, 175–176. 10.1099/ijsem.0.001589

Tortosa, P. & Dubnau, D. (1999) Competence for transformation: A matter of taste. *Current Opinion in Microbiology* **2**: 6, 588–592. https://doi.org/10.1016/S1369-5274(99)00026-0

Tortosa, P., Logsdon, L., Kraigher, B., Itoh, Y., Mandić-Mulec, I. & Dubnau, D. (2001) Specificity and genetic polymorphism of the *Bacillus* competence quorum-sensing system. *Journal of Bacteriology* **183**: 2, 451–460. 10.1128/JB.183.2.451-460.2001

Tran, L. S., Nagai, T. & Itoh, Y. (2000) Divergent structure of the ComQXPA quorum-sensing components: Molecular basis of strain-specific communication mechanism in *Bacillus subtilis*. *Molecular Microbiology* **37**: 5, 1159–1171. 10.1046/j.1365-2958.2000.02069.x

Tredgold, H. R. (2015) Screening for aerobic endospore-forming bacteria as biocontrol agents for powdery mildew disease of cucurbits. M.Sc. Thesis, University of KwaZulu-Natal, Pietermaritzburg, KwaZulu-Natal, South Africa.

Vasantha, N. & Freese, E. (1980) Enzyme changes during *Bacillus subtilis* sporulation caused by deprivation of guanine nucleotides. *Journal of Bacteriology* **144**: 3, 1119–1125.

Williams, P. (2007) Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* **153**: 3923–3938. 10.1099/mic.0.2007/012856-0

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 1, 134–145. 10.1186/1471-2105-13-134

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**:1/2, 203–214. 10.1089/10665270050081478

CHAPTER FIVE

GENERAL OVERVIEW AND CONCLUSIONS

5.1. Summary of findings

Members of the *B. subtilis* group of related taxa play important roles in plant health and growth maintenance, and are celebrated as biocontrol agents and biofertilisers for agriculture (Borriss, 2001, Govindasamy *et al*., 2010). *Bacillus* spp. demonstrate adaptation to specific ecological niches, and these "ecotypes" are functionally adapted to survive the demands of their chosen niche (Mandić-Mulec *et al*., 2015). The formation of biofilms is key to colonisation and persistence of beneficial *B. subtilis sensu lato* populations in the plant-associated environment (Altaf *et al.*, 2017). Biofilm formation is a consequence of social behaviours in *B. subtilis sensu lato*, and aspects of this sociality are governed by the hypervariable ComX peptide (Kalamara *et al*., 2018). ComX variation then serves to protect public goods and ensure discrete communication amongst compatible populations (Stefanic and Mandić-Mulec, 2009). In this way, pherotype variation in *B. subtilis sensu lato* may represent a form—or consequence—of ecological specialisation (Stefanic *et al*., 2012).

Pherotype divergence has been linked to ecological diversification in *B. subtilis* by restricting access to the population's public goods, and an ecotype may demonstrate a predominant ComX variant. By this, pherotype diversity may then be interpreted as an adaptation to ecological diversity in *B. subtilis sensu lato*. Variation in pherotype amongst the *B. subtilis sensu lato* isolates suggested that multiple pherotypes were present within a single ecotype grouping. This has been reported previously by Stefanic *et al.* (2012), with the observation that an ecotype was associated with a dominant pherotype but that there existed minor pherotypes within the population, and that these communicated with the shared pherotypes in other nearby ecotypes. This observation is explained by the "pherotype cycling" hypothesis, wherein the dominant pherotype is constantly evolving (Eldar, 2011, Stefanic *et al*., 2012). The changing of pherotype dominance would link to fitness in an ecotype and avoidance of cheater populations (Eldar, 2011, Smith and Schuster, 2019), ultimately underpinning the ecotype's ability to persist and compete in its target niche.

A research study was embarked upon to explore ecotype and pherotype diversity amongst plant-associated isolates of *B. subtilis* and *B. velezensis*. These isolates had been obtained from the phyllosphere and rhizosphere of several crop species sampled throughout the Midlands area of KwaZulu-Natal, South Africa, and had been screened for their plant-health-promoting characteristics, chiefly surfactin production, and for biocontrol capabilities against fungal phytopathogens. Exploratory studies applied gene sequence data of representative *B. subtilis sensu lato* genomes sourced from the GenBank database, and investigated housekeeping gene candidates to develop a MLSA scheme suited to *B. subtilis sensu lato.* These representative genomes were also examined for sequence variation in the *comQXP* gene region to evaluate pherotype diversity. This gene region was further applied to develop a *comQXP* PCR-RFLP pherotyping protocol using simulated PCR and restriction digests of predicted amplicons. The MLSA scheme was carried forward to application in the isolate subset in order to resolve putative ecotype variation, and this dataset applied to ecotype simulation analysis to corroborate putative ecotype groupings. PCR-based fingerprinting techniques Rep-PCR and RAPD-PCR were evaluated for their potential for rapid ecotyping. Isolate pherotypes were then investigated using the *com*Q gene sequence and *comQXP* PCR-RFLP, and applied a *srfA*-LacZ reporter gene assay using well-known *Bacillus* spp. pherotypes to substantiate putative pherotypes differentiated by the sequence-based approaches.

This study established that:

- Of the nine housekeeping gene targets evaluated for the differentiation of closely-related representative *B. subtilis sensu lato* strains, a concatenation of four housekeeping genes (viz., 16S rRNA, *dnaJ*, *gyrA*, *rpoB*) achieved sufficient resolution of inter- and intra-species diversity. This MLSA was able to resolve sub-cluster variation that was highly suggestive of ecotype variance.

- The in silico *comQ* phylogeny was able to identify several possible pherotype groups amongst reference strains, and, in some strains, determined sequence divergence from known pherotype counterparts. Furthermore, novel putative pherotype clusters were resolved in *B. velezensis*. The in silico *comQXP* amplicon and consequent PCR-RFLP demonstrated four restriction enzyme candidates that had potential for rapid pherotyping: *Bts*CI, *Fnu*4HI, *Cac*8I, and *Hpy*166II. The PCR-RFLP also detected

significant levels of pherotype variation amongst reference strains, and showed divergence in some strains from known pherotypes counterparts. Novel putative pherotype groups were resolved in *B. velezensis*.

- The MLSA of four housekeeping gene targets (viz., 16S rRNA-*dnaJ-gyrA-rpoB*) was able to resolve sub-cluster groupings amongst the subset of isolates that were corroborated to represent putative ecotypes using ES analysis. The ES identified eight putative ecotypes: Six within *B. velezensis* and two within *B. subtilis* isolates. Four of these putative *B. velezensis* ecotypes did not incorporate any of the *B. subtilis sensu lato* reference strains. The PCR fingerprinting techniques failed to corroborate all of the ecotype groupings amongst the isolates, but were able to differentiate between the *B. subtilis* and *B. velezensis* species.

- The isolates demonstrated ES and MLSA groupings sufficiently disparate from reference strains to be suggestive of locally-adapted plant-associated populations. This was particularly evident amongst the *B. velezensis* isolates, which grouped in clades and ecotypes distinct from renowned plant-associated model strain B. *velezensis* FZB42. These ecotype groupings comprised isolates obtained from different crop species and locations, which suggests a degree of localised specialisation and adaptation to prevailing conditions on crops species growing in the KwaZulu-Natal region.

- Phylogenetic analysis of *comQ* sequences from the isolates and known pherotype counterparts determined two distinct clusters each amongst the *B. velezensis* and *B. subtilis* isolates, with only two isolates (viz., bnd134 and bnd156) grouping with known pherotypes. A gene sequence similarity matrix indicated some areas of high levels of sequence similarity which were not observed in the pherotyping assay outcomes.

- The *comQXP* PCR-RFLP was able to corroborate groupings in the *comQ* phylogeny with enzymes *Bts*CI and *Fnu*4HI. Restriction enzyme *Bts*CI demonstrated sensitivity towards *B. subtilis* (in particular the isolates comprising the *Bacillus* sp. JS MLSA clade), while restriction enzyme *Fnu*4HI proved able to differentiate between more closely-related groupings in *B. velezensis*. This demonstrates that *comQXP* PCR-RFLP

has potential for application as a pherotype screening technique for beneficial *B. subtilis sensu lato* isolates.

- Simulated *comQXP* PCR-RFLP identified four enzyme candidates suited to distinguishing between pherotypes amongst representative *B. subtilis* sensu lato. It was unexpected that *Hpy*166II would display fingerprints highly similar to those of *Bts*CI in the isolate *comQXP*, as this was not indicated in the simulated PCR-RFLP. The only reference strain which was able to be applied in the isolate *comQXP* PCR-RFLP was *B. subtilis* subsp. *subtilis* DSM 10^{T}. This reference strain shows very similar fingerprints to those projected for *B. subtilis* in the simulations for *Bts*CI and *Fnu*4HI. In silico PCR-RFLP is therefore reasonably accurate for the purposes of identifying PCR-RFLP restriction enzyme candidates.

- The pherotyping *srfA*-LacZ reporter gene assay demonstrated significantly higher level of pherotype reactivity amongst the *B. velezensis* isolates, but failed to support the pherotype groupings defined by the *comQ* phylogeny. The *B. subtilis* isolates achieved lower levels of tester strain responses than did the *B. velezensis* isolates. Some isolates did not return any matches to the pherotype strains tested. Most of the *B. velezensis* isolates matched to multiple pherotype strains, with three pherotypes in predominating (viz., 168, RO-C-2, and RO-H-1). This assay requires further optimisation to better accommodate *B. velezensis* isolates.

- An MLSA clade comprising isolates B81, CT-R67, bnd136, sqo221, and sqo271 clustered closely with reference strain Bacillus sp. JS. This cluster was distinct to the reference strains most closely-related to the *B. subtilis sensu stricto* in the MLSA and was confirmed by ES to constitute a single putative ecotype. Furthermore, this was the only putative ecotype grouping discriminated as a common profile by fingerprinting methods Rep and RAPD OPG-16. This grouping of isolates also comprised a single clade in the comQ phylogeny, indicating a single pherotype unique to this grouping. That these isolates clustered with gene sequences from a Korean, plant-associated strain of biocontrol potential was interesting, as it highlighted the possibility of ecotype divergence and pherotype conservation spanning continents. Bacillus sp. JS is also taxonomically interesting as it formed a distinct MLSA clade within the taxa most

closely-related to *B. subtilis*. Moreover, these isolates originated from various crop types and sampling locations, evidencing a form of localised adaptation. The group also formed a distinct pherotype clade in the *comQ* phylogeny and *comQXP* PCR-RFLP with *Bts*CI, but no similarities could be drawn regarding tester strain stimulation in the pherotyping assay.

- The putative ecotype groupings amongst isolates defined by ES did not correlate entirely with the clusters defined in the *comQ* phylogeny. This indicates that there may be multiple pherotypes within a single ecotype grouping.

- The *B. velezensis* isolates demonstrated the greatest levels of distinctness from representative genome reference strains and pherotypes across all investigations. These isolates formed unique ecotype and pherotype groupings that were frequently distinct from the *B. velezensis* model strain FZB42, and to other representative species counterparts. This indicates that these isolates had undergone substantial specialisation in terms of ecotype and pherotype in order to meet the demands of the plant-associated environment.

5.2. Future research prospects

This research may be expanded in future studies to include several aspects of ecotype and pherotype divergence in plant-associated *B. subtilis sensu lato*:

- Isolates of *Bacillus* spp. have been reported to demonstrate kin discrimination behaviours. Stefanic *et al.* (2015) observed that soil-derived isolates of *Bacillus* spp. recognised their most-closely related counterpart isolates, and displayed co-operative swarming behaviours. It would be of interest to determine whether such kin discrimination behaviours occur within and between the plant-associated ecotypes applied in this study. Kin discrimination and recognition also represent a form of sociality in *Bacillus* spp. that could contribute to ecological niche fitness. No studies of kin discrimination have been carried out in *B. velezensis*, and this represents an

attractive research opportunity in the context of beneficial strain colonisation and establishment on the plant host, and possible synergy between related populations.

- Chemical characterisation of isolate ComX variants would contribute greatly to our appreciation of divergent plant-associated ComX variants, and may enable more quantifiable and standardised conditions for the reporter gene assay in order to determine isolate-ComX relationships to known pherotype strains.

- The resolution of inter- and intra-strain relationships afforded by fingerprinting PCR deserves further exploration for suitability as rapid-ecotyping methods. There are a wealth of fingerprinting primer sets not evaluated within the context of this study and that may hold suitable ecotyping candidates.

- The OPG-11 fingerprinting profile for *B. velezensis* isolates identified a common band approximately 2000 bp in size, which may hold some future potential as a marker for plant-associated *B. velezensis*.

- The relationships of the five isolates related to *Bacillus* sp. JS present an enticing prospect for further study. That isolates formed MLSA and ecotype relationships to this particular strain—one of reported biocontrol importance—brings to the fore many research possibilities as to these isolates' colonisation abilities and biocontrol capabilities of these isolates, and the importance of this group as a possibly globally-represented plant-associated ecotype.

5.3. Conclusions

Ecological specialisation to suit the prevailing conditions of a target niche has been described in many bacterial populations (Cohan, 2017). These ecotypes demonstrate functional variations to other members of the same species which enable them to colonise and persist within their target niche (Lan and Reeves, 2000). In beneficial *Bacillus* spp. existing in association with plants, ecological specialisation may hold the key to our understanding of the characteristics that allow the colonisation of plant-associated niches and which underpin the persistence of these beneficial populations. Ecotype groupings amongst isolates were defined in this study

using a MLSA approach, and these putative ecotype groupings provided evidence for local adaptation. This is an important finding for such endeavours as biocontrol agent screening, where the ability of an introduced strain to compete and establish within the prevailing conditions of the target niche is key to biocontrol strategy success.

An important aspect of bacterial colonisation of the plant host is the formation of biofilms. Social behaviours like biofilm formation in *B. subtilis sensu lato* are governed by the hypervariable ComX peptide, resulting in localised communication restricted to members of a ComX variant (i.e. pherotype). The *comQ* phylogeny demonstrated pherotype groupings amongst isolates of the same species—in particular *B. velezensis*—which contained representatives from different environments, plant parts, and crop species. This could further suggest a regional influence or predominant pherotype class at play amongst plant-associated populations. The fact that the phylogeny demonstrated that the pherotypes inferred by *comQ* relatedness bore some similarity to each other—albeit with evidence of divergence in some clades—suggests that there may a limited pool of pherotypes available across the *B. subtilis sensu lato*. That other studies have been able to detect pherotypes in common to the "known" pherotype subset in isolates from soil and rhizoplane environments seems to agree with this (Stefanic *et al.*, 2012, Oslizlo *et al.*, 2015). It could be postulated that some pherotypes may be cross-reactive, as was observed in the pherotype assay, and that ecotype populations may select a pherotype that benefits them within the context of their region. These benefits could arise from a common goods perspective, or in allowing these strains to become more promiscuous and derive benefit from other populations reacting to their quorum-sensing signals.

The sequence-based investigations of pherotype conducted in this study found evidence of shared and somewhat-related pherotypes which may suggest that the ComQXPA machinery is transferable between species and populations, and this concept has been reported in other studies (Tortosa and Dubnau, 1999, Ansaldi and Dubnau, 2004, Eldar, 2011). This could be taken a step further to postulate that dominant pherotypes may be altered according to the needs of the population or the predominance of a particular pherotype(s) in that environment. The pherotype cycling hypothesis speaks to this, with the underlying reasoning being that pherotypes evolve as a means of circumventing cheating populations within a community of

related organisms. In a highly-competitive environment, such as that found on the plant, this would hold true. *Bacillus* spp. are ubiquitous in the plant- and soil-associated environment and it is in the best interest of populations to maintain their ecological fitness by protecting public goods from exploitation by non-public-goods-producing competitors. The concept of pherotype dominance and cycling amongst plant-associated ecotypes specifically, and the implications of these phenomena in biocontrol strategies, particularly amongst *B. velezensis*, remains to be elucidated.

This study reflects trends observed at Stefanic *et al.* (2012) with regards to ecotype groups comprising major and minor pherotypes within them, it is of interest to explore the range of plant-associated *B. subtilis sensu lato* pherotypes—particularly amongst ecotypes of the unexplored *B. velezensis* species—in an effort to better understand the roles and importance of ecological specialisation and communication amongst beneficial strains for use in biofertiliser and biocontrol applications. The *B. subtilis* and *B. velezensis* evaluated in this study demonstrated a degree of local adaption in terms of distinct ecotype groupings and pherotype sequences distantly related to those of "known" *B. subtilis sensu lato* pherotype variants. That these isolates also represent candidates for use in biocontrol applications presents enticing prospects for further study of the ecological fitness of localised ecotypes and their prevalence in the agricultural context.

5.4. References

Altaf, M. M., A., I., Khan , M. S. A. & Grohmann, E. (2017) *Bacillus* biofilms and their role in plant health. *Biofilms in Plant and Soil Health* (Ahmad, I. & Husain, F. M., eds.), pp. 55–67. https://doi.org/10.1002/9781119246329.ch4

Ansaldi, M. & Dubnau, D. (2004) Diversifying selection at the *Bacillus* quorum-sensing locus and determinants of modification specificity during synthesis of the ComX pheromone. *Journal of Bacteriology* **186**: 1, 15–21. 10.1128/jb.186.1.15-21.2004

Borriss, R. (2011) Use of plant-associated *Bacillus* strains as biofertilizers and biocontrol agents in agriculture. *Bacteria in Agrobiology: Plant Growth Responses* (Maheshwari, D. K., ed.) pp. 41–76. Springer-Verlag, Berlin, Germany. 10.1007/978-3-642-20332-9_3

Cohan, F. M. (2017) Transmission in the origins of bacterial diversity, from ecotypes to phyla. *Microbiology Spectrum* **5**: 5, 1–26. doi:10.1128/microbiolspec.MTBP-0014-2016

Eldar, A. (2011) Social conflict drives the evolutionary divergence of quorum sensing. *Proceedings of the National Academy of Sciences Of the United States of America* **108**: 33, 13635–13640. 10.1073/pnas.1102923108

Govindasamy, V., Senthilkumar, M., Magheshwaran, V., Kumar, U., Bose, P., Sharma, V. & Annapurna, K. (2010) *Bacillus* and *Paenibacillus* spp.: Potential PGPR for sustainable agriculture. *Plant Growth and Health Promoting Bacteria* (Maheshwari, D. K., ed.) pp. 333–364. Springer-Verlag, Berlin, Germany. https://doi.org/10.1007/978-3-642-13612-2

Kalamara, M., Spacapan, M., Mandić-Mulec, I. & Stanley-Wall, N. R. (2018) Social behaviours by *Bacillus subtilis*: Quorum sensing, kin discrimination and beyond. **110**: 6, 863–878. 10.1111/mmi.14127

Lan, R. & Reeves, P. R. (2000) Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends in Microbiology* **8**: 9, 396–401. 10.1016/S0966-842X(00)01791-1

Mandić-Mulec, I., Stefanic, P. & van Elsas, J. D. (2015) Ecology of *Bacillaceae*. *Microbiology Spectrum* **3**: 1, 1–23. 10.1128/microbiolspec.TBS-0017-2013

Oslizlo, A., Stefanic, P., Vatovec, S., Beigot Glaser, S., Rupnik, M. & Mandić-Mulec, I. (2015) Exploring ComQXPA quorum-sensing diversity and biocontrol potential of *Bacillus* spp. isolates from tomato rhizoplane. *Microbial Biotechnology* **8**: 3, 527–540. 10.1111/1751-7915.12258

Smith, P. & Schuster, M. (2019) Public goods and cheating in microbes. *Current Biology* **29**: 11, R442–R447. https://doi.org/10.1016/j.cub.2019.03.001

Stefanic, P. & Mandić-Mulec, I. (2009) Social interactions and distribution of *Bacillus subtilis* pherotypes at microscale. *Journal of Bacteriology* **191**: 6, 1756–1764. 10.1128/jb.01290-08

Stefanic, P., Kraigher, B., Lyons, N. A., Kolter, R. & Mandic-Mulec, I. (2015) Kin discrimination between sympatric *Bacillus subtilis* isolates. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 45, 14042-14047. 10.1073/pnas.1512671112

Stefanic, P., Decorosi, F., Viti, C., Petito, J., Cohan, F. M. & Mandić-Mulec, I. (2012) The quorum sensing diversity within and between ecotypes of *Bacillus subtilis*. *Environmental Microbiology* **14**: 6, 1378–1389. 10.1111/j.1462-2920.2012.02717.x

Tortosa, P. & Dubnau, D. (1999) Competence for transformation: A matter of taste. *Current Opinion in Microbiology* **2**: 6, 588–592. https://doi.org/10.1016/S1369-5274(99)00026-0

APPENDIX A

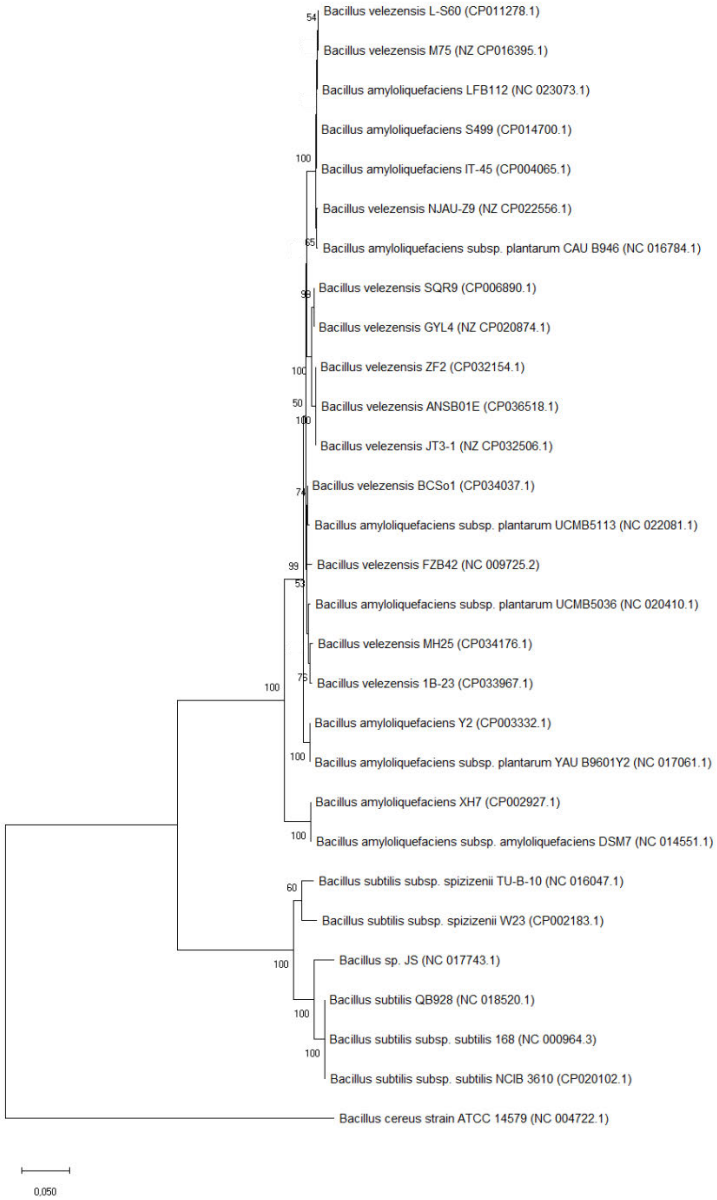


Figure A1. Phylogeny of *cheA* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-8931.52) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 2140 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

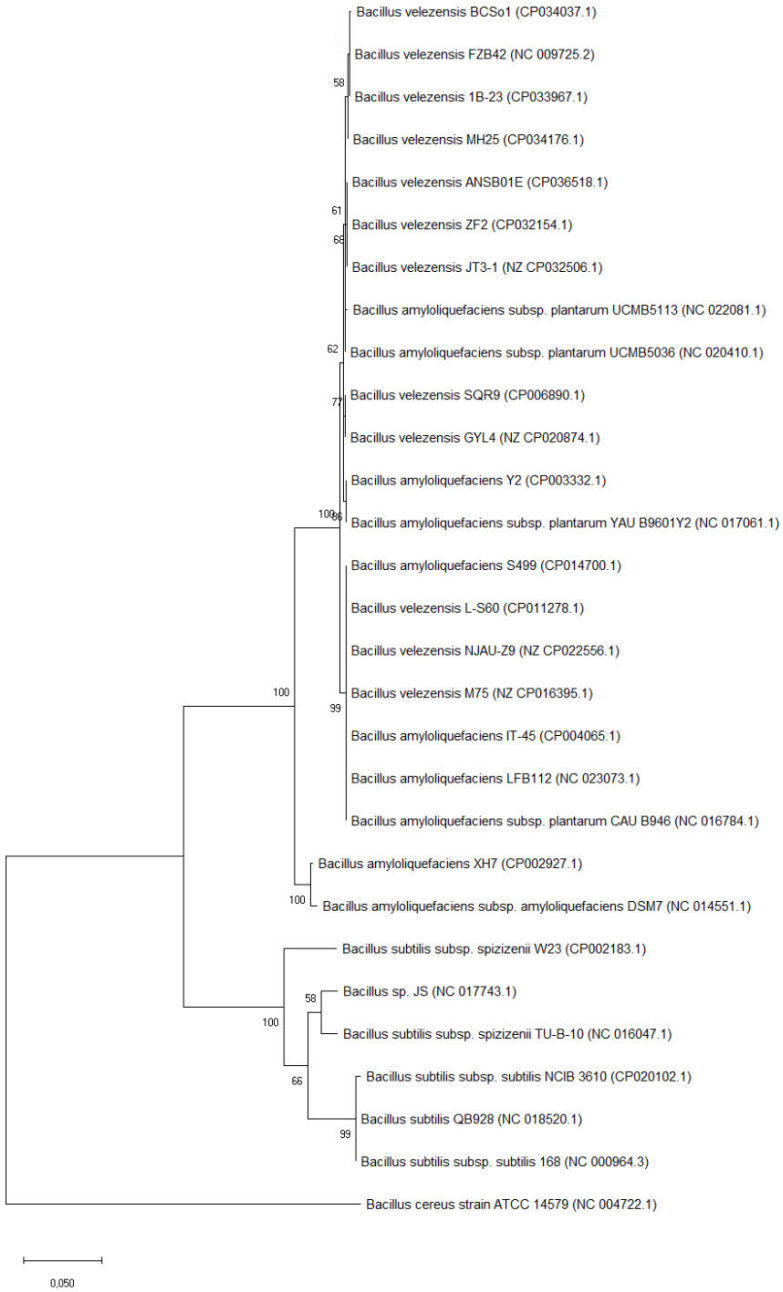


Figure A2. Phylogeny of *dnaJ* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-4243.14) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 1138 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

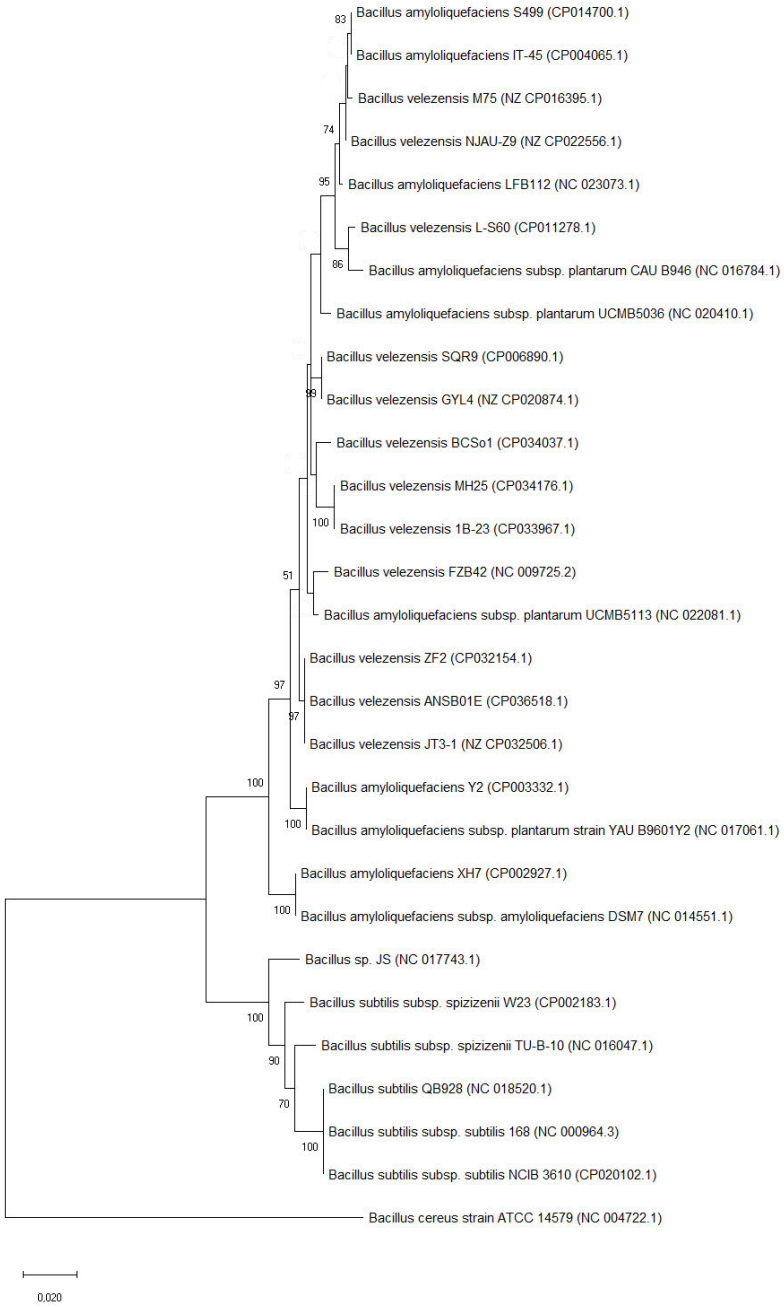


Figure A3. Phylogeny of *groEL* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-5416.79) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 1641 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

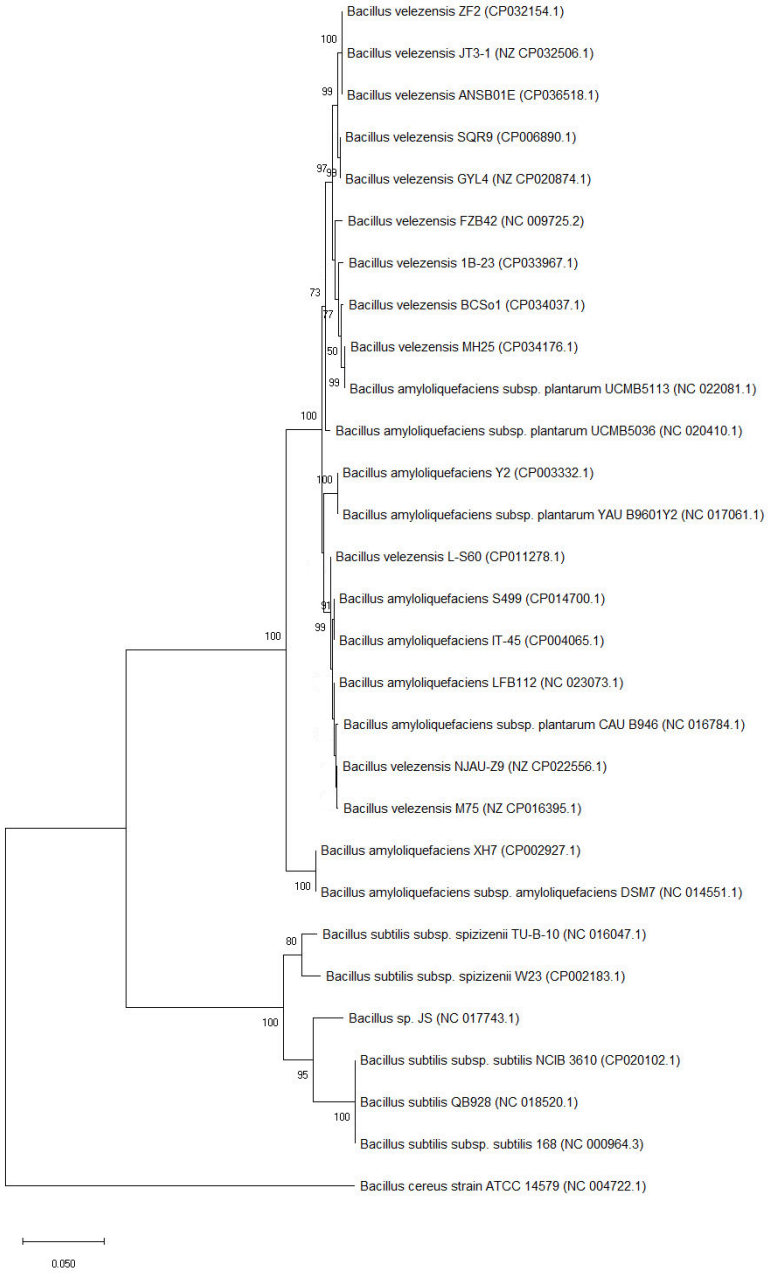


Figure A4. Phylogeny of *gyrA* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-9896.62) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 2466 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

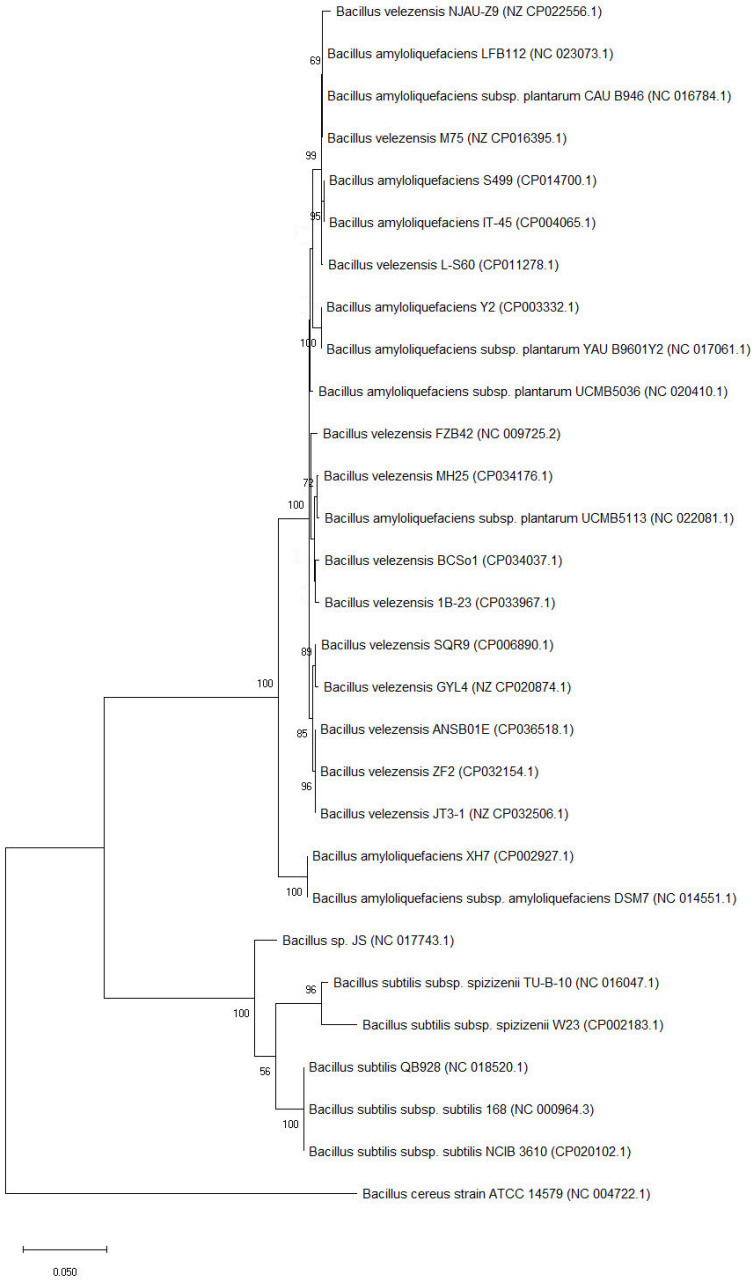


Figure A5. Phylogeny of *gyrB* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-7417.69) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 1903 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

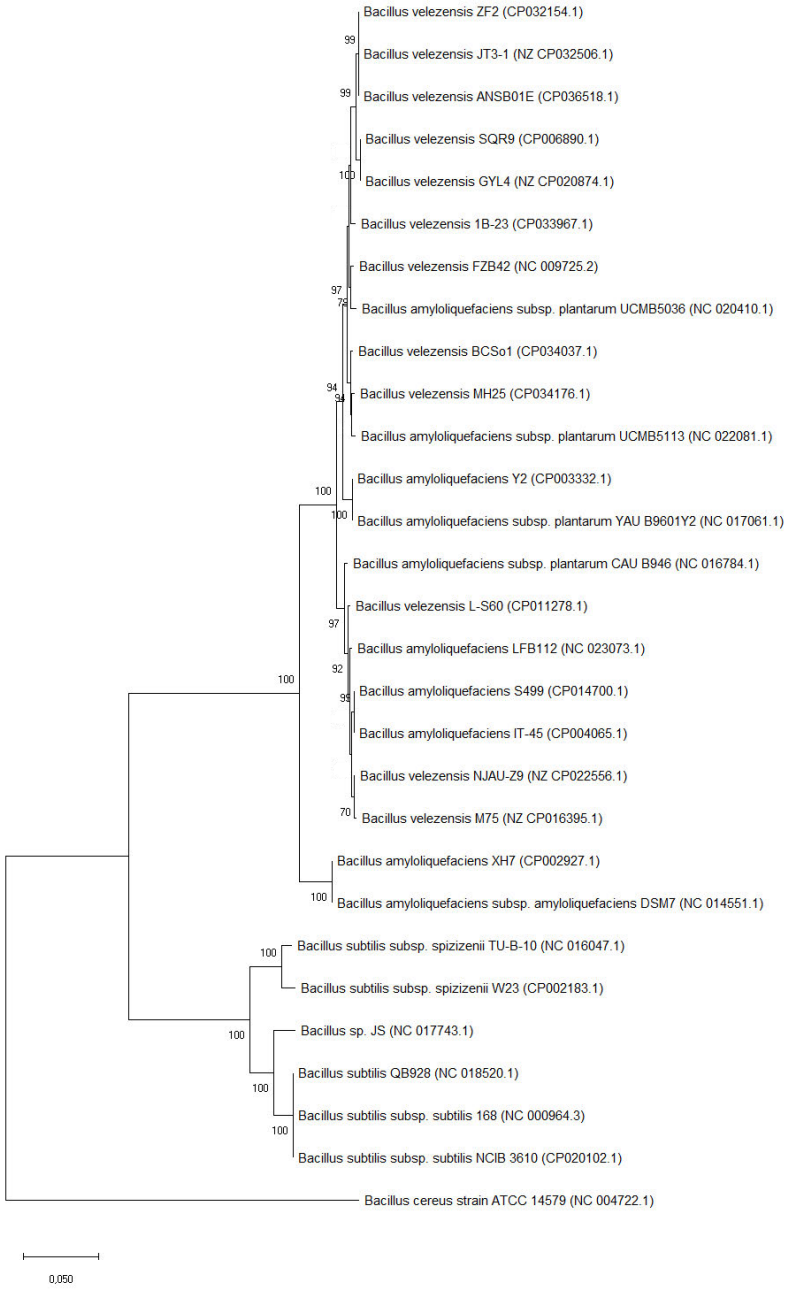


Figure A6. Phylogeny of *polC* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-17517.04) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 4334 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

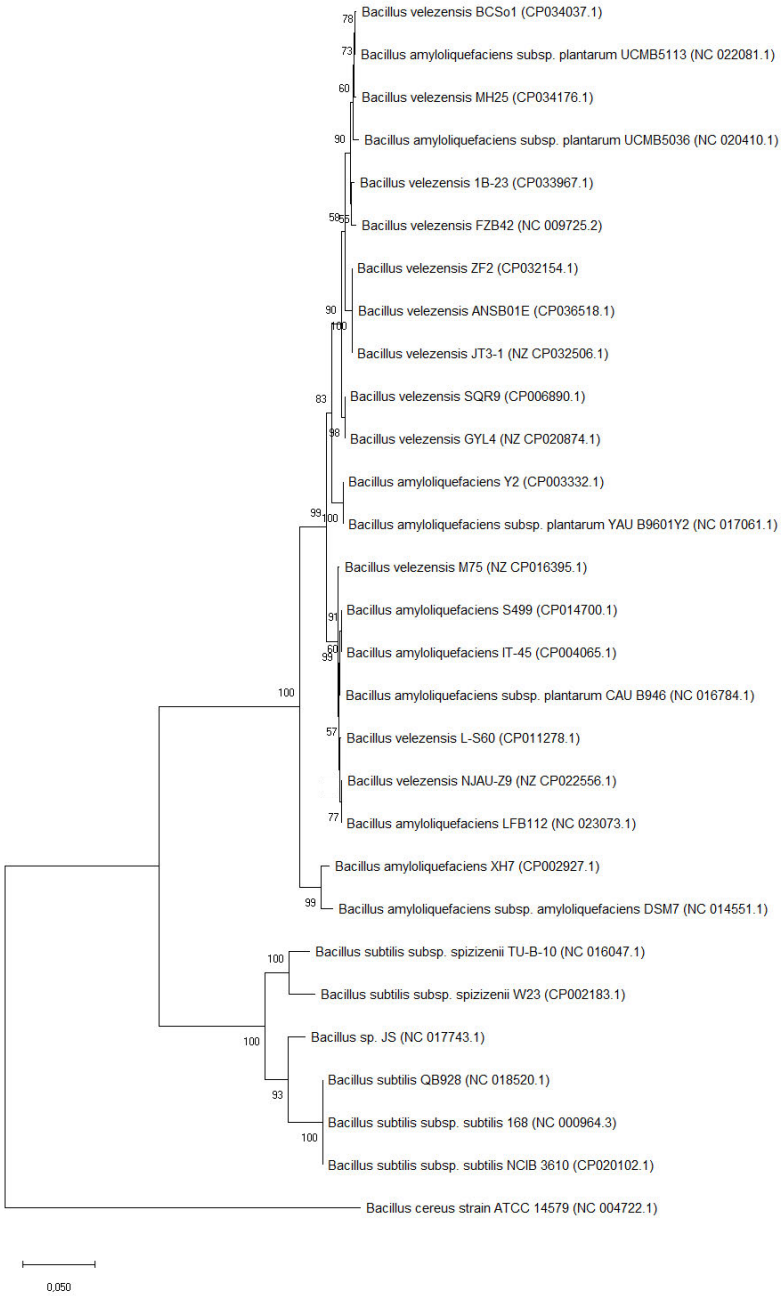


Figure A7. Phylogeny of *purH* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-6461.84) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 1542 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

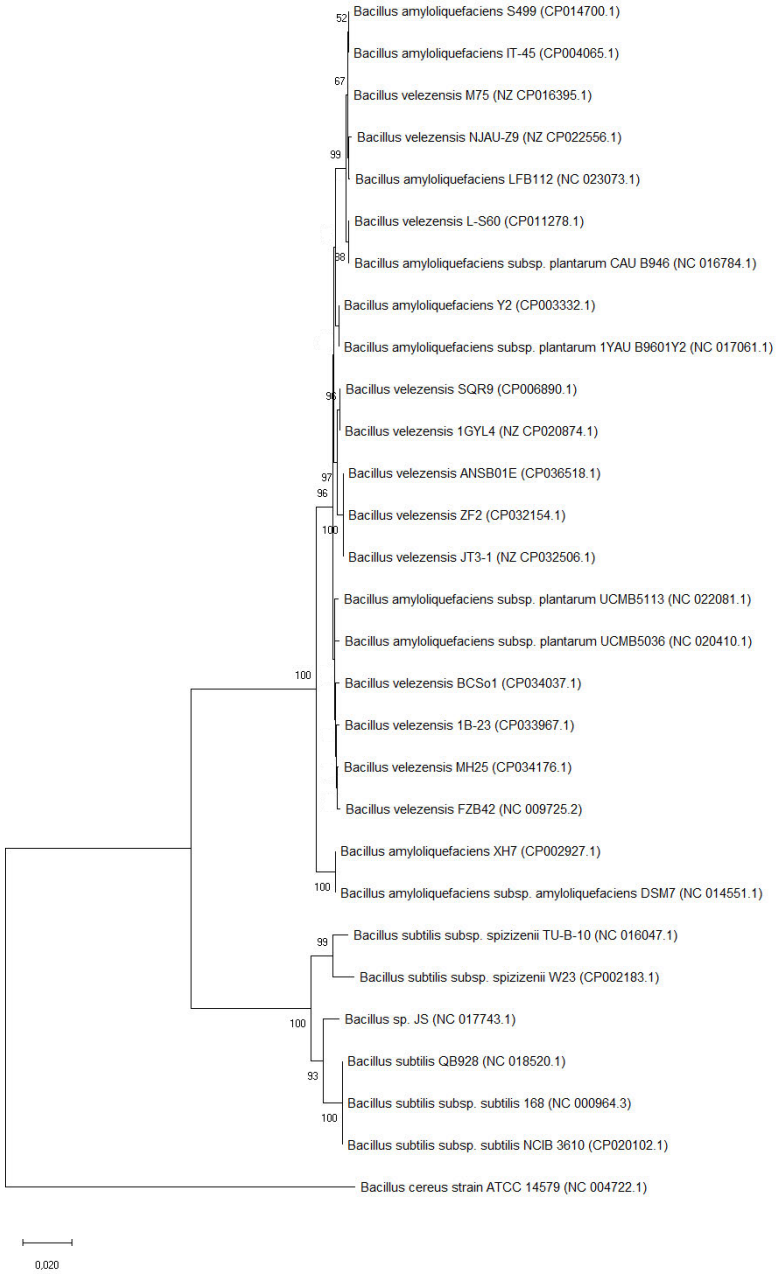


Figure A8. Phylogeny of *rpoB* gene sequences for selected strains of *Bacillus subtilis sensu lato*. The phylogenetic tree was generated in MEGA X and was inferred by the Maximum Likelihood method and Kimura 2-parameter model. This tree reached the highest log likelihood (-10081.19) is shown. The figures next to the branches indicate the percentage of trees in which the associated taxa clustered together. The tree branch lengths are represented in the number of substitutions per site. This analysis applied 29 nucleotide sequences with 3247 positions in the final dataset. Tree branch lengths were measured in the number of substitutions per site, as indicated by the scale bar. Shown next to the branches is the percentage of trees in which the associated taxa clustered together.

APPENDIX B

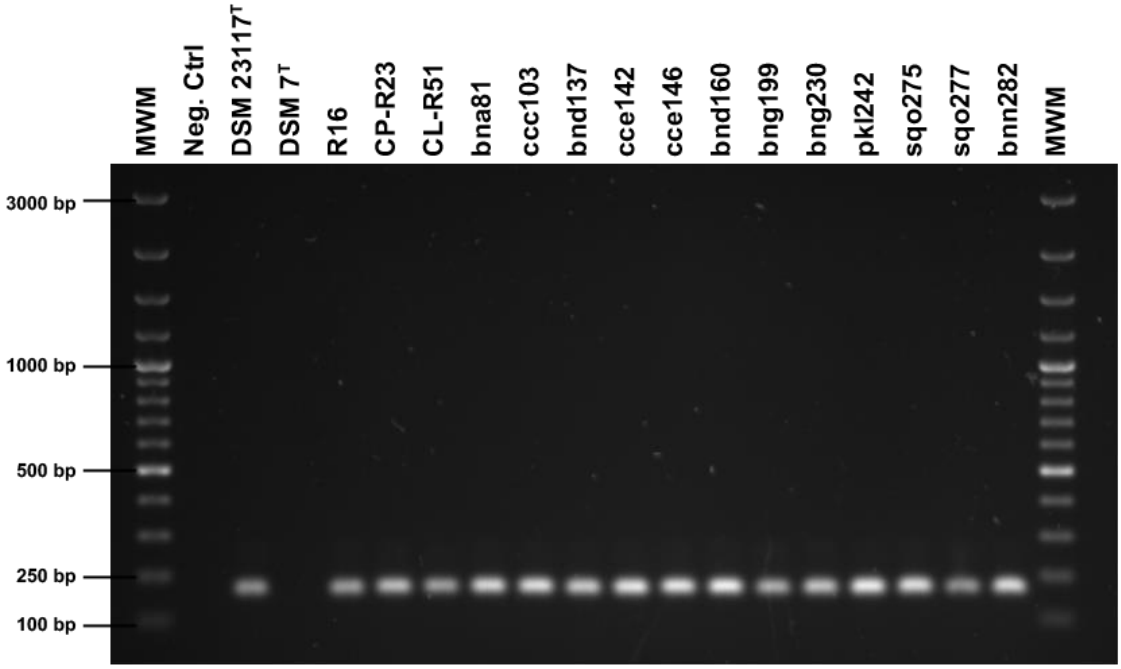


Figure B1. Agarose gel image of *Bacillus velezensis* endpoint PCR of plant-associated putative *B. velezensis* isolates, and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T} and *B. velezensis* DSM 23117^{T}. Isolates belonging to Bacillus velezensis show a ~170 bp fragment targeting a sugar kinase unique to this species (Dunlap, 2019). Agarose gel comprised 1 % w/v and used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Condalab, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".

Table B1. Similarity matches for partial gyrase subunit alpha sequences of *Bacillus subtilis sensu lato* isolates after a BLAST search of the NCBI GenBank BLASTN databases (Last Accessed: May 2020). Search parameters were restricted to within the nr/nt database, limited to the *B. subtilis* group (Taxid: 653685), and excluded uncultured/environmental samples.

Isolate	Best BLAST Database Match	GenBank Accession No.	Max Score	Total Score	Query Cover	E-value	Per. Ident
bna81	Bacillus velezensis strain ZeaDK315Endobac16 chromosome	CP043809.1	1676	1676	100%	0	99.78%
bnd137	Bacillus velezensis strain WLYS23 chromosome, complete genome	CP055160.1	1666	1666	100%	0	99.67%
bnd160	Bacillus velezensis strain ZeaDK315Endobac16 chromosome	CP043809.1	1692	1692	100%	0	100.00%
bng199	Bacillus velezensis strain B268 chromosome, complete genome	CP053764.1	1692	1692	100%	0	100.00%
bng230	Bacillus velezensis strain WLYS23 chromosome, complete genome	CP055160.1	1676	1676	100%	0	99.67%
bnn282	Bacillus velezensis strain BY6 chromosome, complete genome	CP051011.1	1692	1692	100%	0	100.00%
cce142	Bacillus velezensis strain WLYS23 chromosome, complete genome	CP055160.1	1674	1674	100%	0	99.67%
cce146	Bacillus velezensis strain WLYS23 chromosome, complete genome	CP055160.1	1690	1690	100%	0	100.00%
ccc103	Bacillus velezensis strain ZeaDK315Endobac16 chromosome	CP043809.1	1692	1692	100%	0	99.89%
pkl242	Bacillus velezensis strain WLYS23 chromosome, complete genome	CP055160.1	1676	1676	99%	0	99.67%
sqo277	Bacillus velezensis strain 8-2 chromosome, complete genome	CP028439.1	1692	1692	100%	0	100.00%
sqo275	Bacillus velezensis strain WLYS23 chromosome, complete genome	CP055160.1	1672	1672	100%	0	99.67%
bnd136	Bacillus subtilis strain BJ3-2 chromosome, complete genome	CP025941.1	1652	1652	100%	0	99.34%
bnd134	Bacillus subtilis strain PR10 chromosome, complete genome	CP040528.1	1687	1687	99%	0	99.89%
bnd156	Bacillus subtilis strain SRCM103571 chromosome, complete genome	CP035231.1	1650	1650	100%	0	99.56%
bng215	Bacillus subtilis strain TLO3, complete genome	CP021169.1	1688	1688	100%	0	99.78%
bng221	Bacillus subtilis strain BJ3-2 chromosome, complete genome	CP025941.1	1659	1659	99%	0	99.34%
sqo271	Bacillus subtilis strain BJ3-2 chromosome, complete genome	CP025941.1	1659	1659	99%	0	99.34%
sqo279	Bacillus subtilis strain SRCM102754 chromosome, complete genome	CP028202.1	1668	1668	100%	0	99.89%

Table B2. PCR reaction conditions for the primers used in DNA fingerprinting and housekeeping gene amplification.

Protocol	Initialisation	PCR cycling (Temperature and duration)			Cycles	Final Extension
		Denaturation	Annealing	Elongation		
Rep-PCR	95°C / 2 min	95°C / 30 sec	45°C / 30 sec	72°C / 1 min	35	72°C / 5 min
RAPD-PCR	95°C / 2 min	95°C / 30 sec	36°C / 30 sec	72°C / 1 min	35	72°C / 5 min
16S rRNA*	94°C / 5 min	94°C / 1 min	65°C / 90 sec	72°C / 2 min	35	72°C / 10 min
dnaJ	95°C / 2 min	95°C / 30 sec	58°C / 30 sec	72°C / 1 min	30	72°C / 5 min
*gyrA**	95°C / 2 min	95°C / 1 min	60°C / 30 sec	72°C / 1 min	30	72°C / 5 min
rpoB	95°C / 2 min	95°C / 30 sec	48°C / 30 sec	72°C / 1 min	30	72°C / 5 min

*Amplicons for 16S rRNA and *gyrA* were obtained in previous studies (Tredgold, 2015, Hunter, 2016). These PCRs using Promega GoTaq PCR reagents (Promega, Madison, Wisconsin, U.S.A.) and each 25 µL PCR reaction consisted of the following: 0.4 µM of the appropriate primers, 1x GoTaq Flexi Buffer (without $MgCl_2$), 200 µM of each deoxyribonucleotide (dNTP), 2.5 U GoTaq polymerase, 1.5 mM $MgCl_2$, 1 µL template DNA, and sufficient nuclease-free water (Promega, Wisconsin, U.S.A.) to bring the final volume to 25 µL.

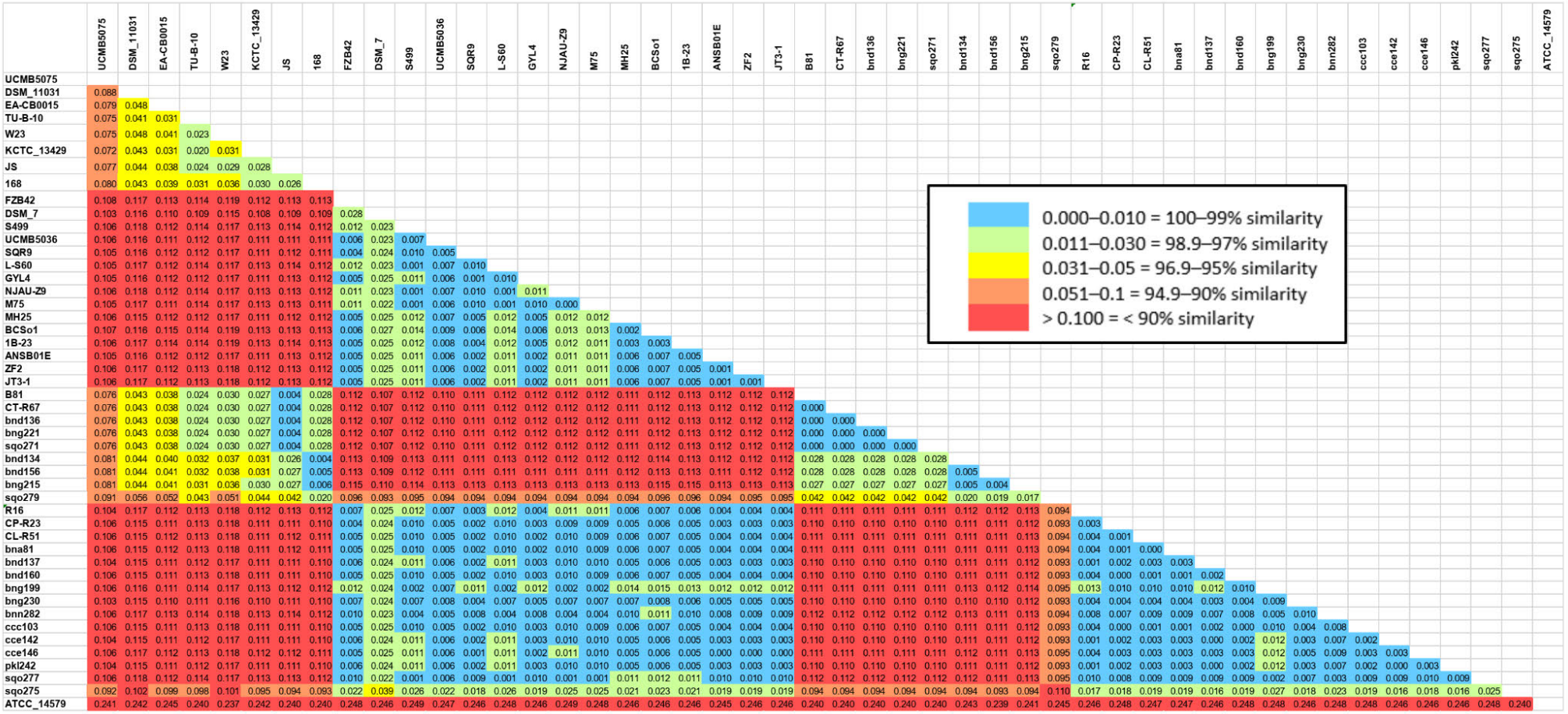
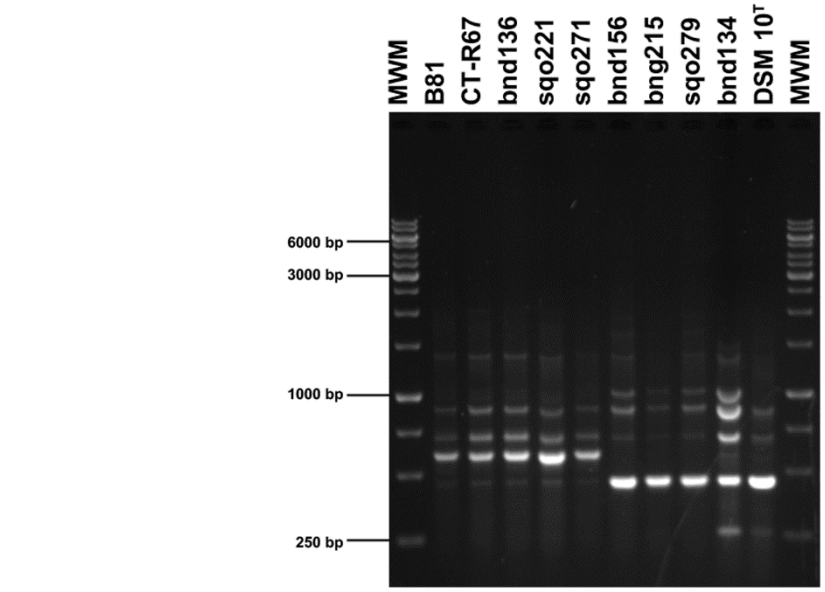
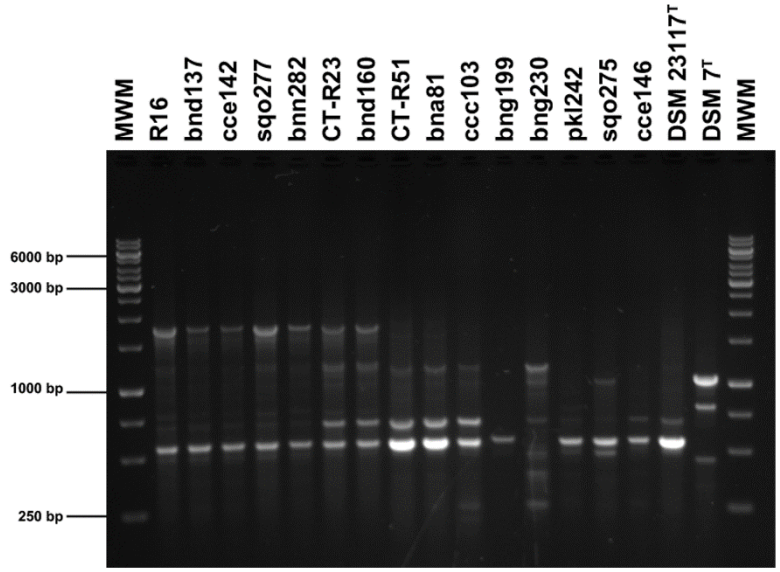


Figure B2. Estimates of evolutionary divergence by a pairwise distance matrix of concatenated MLSA gene sequences for *Bacillus subtilis sensu lato* isolates and reference strains. The matrix was generated in MEGA X (Kumar *et al.*, 2018) using the Maximum Composite Likelihood model . The dataset comprised 48 sequences with a total of 2839 positions. All positions with less than 95% site coverage were eliminated. Figures indicate the number of base substitutions per site between sequences, heatmap colour grading indicates percentage similarity between sequences as shown in the figure key.



(A) *Bacillus subtilis*
Reference strain *Bacillus subtilis* subsp. *subtilis* DSM 10^T



(B) *Bacillus velezensis*
Reference strain *Bacillus amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T and *Bacillus velezensis* DSM 23117^T

Figure B3. Agarose gel images of Rep-PCR fingerprint profiles of plant-associated *Bacillus subtilis sensu lato* isolates, and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T, *B. subtilis* subsp. *subtilis* DSM 10^T, and *B. velezensis* DSM 23117^T. Agarose gel comprised 1.2% w/v and used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Condalab, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".

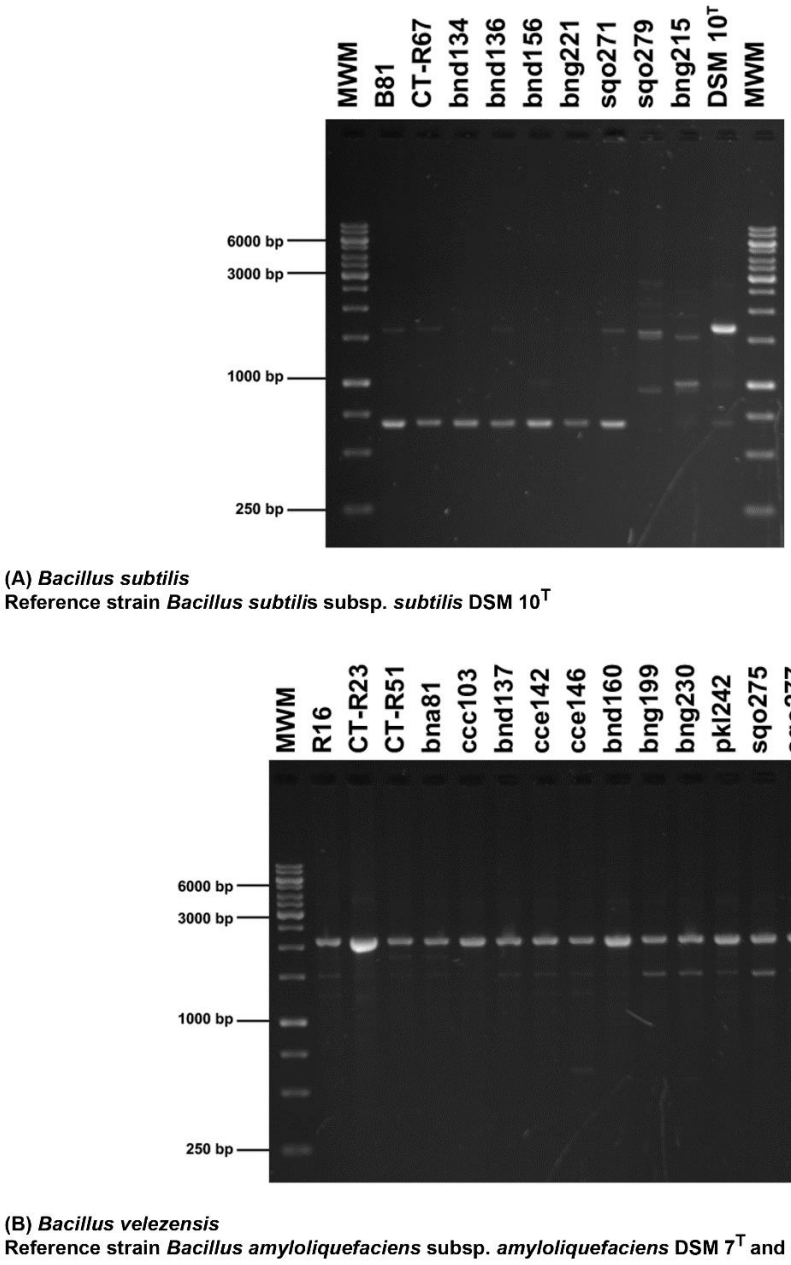


Figure B4. Agarose gel images of RAPD-PCR fingerprint profiles generated using the OPG-11 primer set for plant-associated *Bacillus subtilis sensu lato* isolates, and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^{T}, *B. subtilis* subsp. *subtilis* DSM 10^{T}, and *B. velezensis* DSM 23117^{T}. Agarose gel comprised 1.2% w/v and used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Condalab, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".

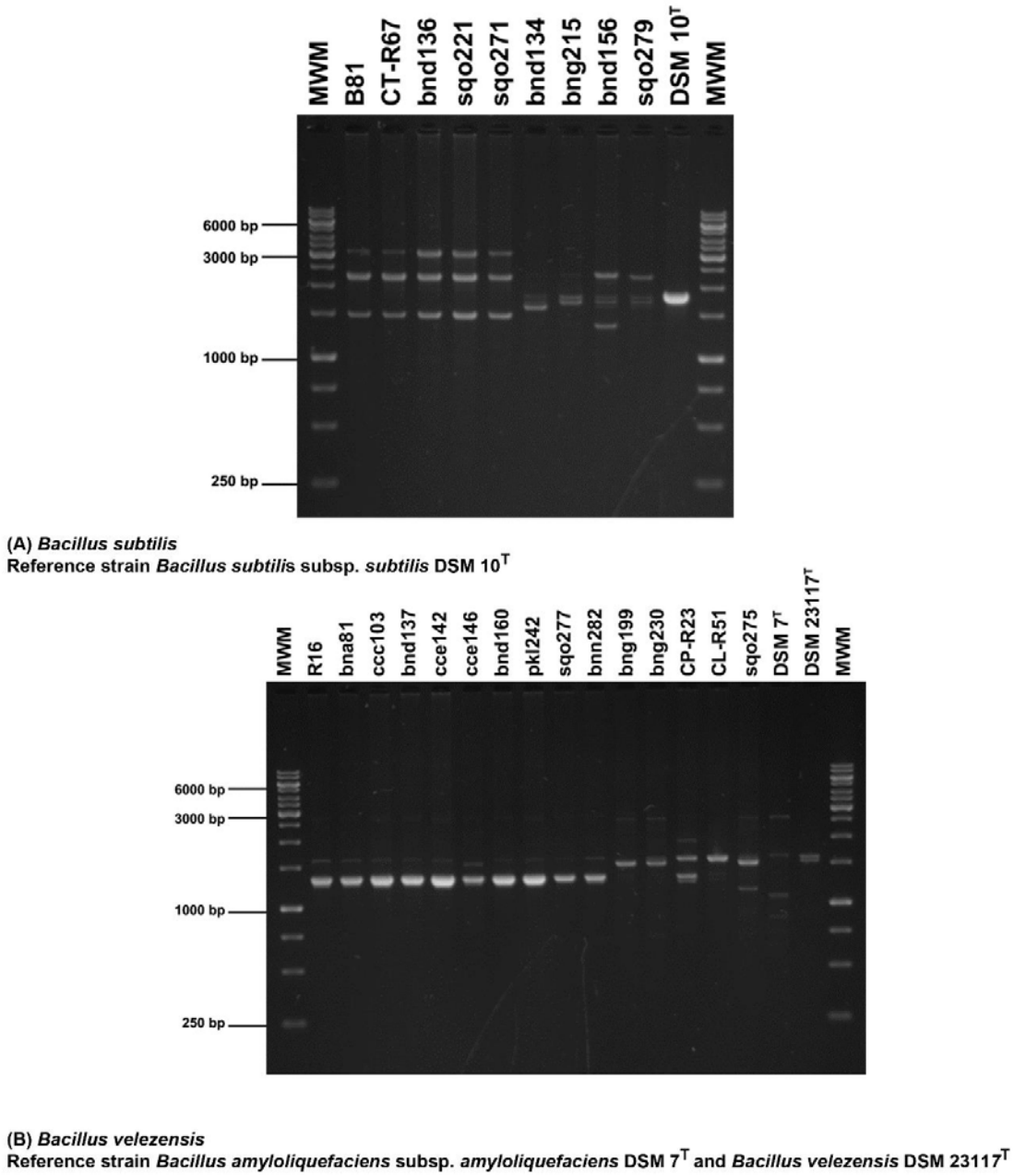


Figure B5. Agarose gel images of RAPD-PCR fingerprint profiles generated using the OPG-16 primer set for plant-associated *Bacillus subtilis sensu lato* isolates, and reference type strains *B. amyloliquefaciens* subsp. *amyloliquefaciens* DSM 7^T, *B. subtilis* subsp. *subtilis* DSM 10^T, and *B. velezensis* DSM 23117^T. Agarose gel comprised 1.2% w/v and used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Condalab, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".

APPENDIX C

Table C1. Similarity matches for partial *comQ* sequences of *B. subtilis sensu lato* isolates using a BLAST search of the NCBI GenBank BLASTN database (Last Accessed: December 2020). Search parameters were restricted to within the nr/nt database, limited to the *B. subtilis* group (Taxid: 653685), and excluded uncultured/environmental samples. E-value for all searches was zero.

Isolate	Top 3 BLAST Matches	GenBank Accession No.	Max / Total Score	Query Cover	Per. Ident
sqo279	*B. subtilis* strain H1, complete genome	CP026662.1	1103	100%	99.83%
	B. subtilis strain SRCM102754, complete genome	CP028202.1	1103	100%	99.83%
	B. subtilis strain PS-160, partial quorum-sensing cds	FJ172581.2	1103	100%	99.83%
bng215	*B. subtilis* strain H1, complete genome	CP026662.1	1411	100%	98.99%
	B. subtilis strain SRCM102754, complete genome	CP028202.1	1411	100%	98.99%
	B. subtilis strain PS-160, partial quorum-sensing cds	FJ172581.2	1411	100%	98.99%
bnd156	*B. subtilis* strain SRCM103689, complete genome	CP035391.1	1332	100%	98.55%
	B. subtilis strain SRCM103571, complete genome	CP035231.1	1332	100%	98.55%
	B. subtilis strain SRCM103612, complete genome	CP035406.1	1327	100%	98.42%
bnd134	*B. subtilis* strain SRCM102745, complete genome	CP028209.1	1297	100%	99.31%
	B. subtilis strain P5_B1, complete genome	CP045817.1	1297	100%	99.31%
	B. subtilis subsp. *subtilis* strain G7, complete genome	CP029609.1	1297	100%	99.31%
CT-R67	*B. subtilis* strain KKD1, complete genome	CP054584.1	719	99%	92.14%
	B. halotolerans strain F41-3, complete genome	CP041357.1	719	99%	92.14%
	B. halotolerans strain XH-1, complete genome	CP061284.1	719	99%	92.14%
bng221	*B. subtilis* strain KKD1, complete genome	CP054584.1	1616	100%	98.07%
	B. halotolerans strain F41-3, complete genome	CP041357.1	1616	100%	98.07%
	B. halotolerans strain XH-1, complete genome	CP061284.1	1616	100%	98.07%

Table C1. Continued.

Isolate	Top 3 BLAST Matches	GenBank Accession No.	Max / Total Score	Query Cover	Per. Ident
bnd136	*B. subtilis* strain KKD1, complete genome	CP054584.1	1229	99%	97.00%
	B. halotolerans strain F41-3, complete genome	CP041357.1	1229	99%	97.00%
	B. halotolerans strain XH-1, complete genome	CP061284.1	1229	99%	97.00%
sqo271	*B. subtilis* strain KKD1, complete genome	CP054584.1	1567	100%	98.22%
	B. halotolerans strain F41-3, complete genome	CP041357.1	1567	100%	98.22%
	B. halotolerans strain XH-1, complete genome	CP061284.1	1567	100%	98.22%
B81	*B. subtilis* strain KKD1, complete genome	CP054584.1	1594	100%	97.94%
	B. halotolerans strain F41-3, complete genome	CP041357.1	1594	100%	97.94%
	B. halotolerans strain XH-1, complete genome	CP061284.1	1594	100%	97.94%
bnd137	*B. amyloliquefaciens* strain KC41, complete genome	CP044444.1	1664	100%	99.56%
	B. velezensis strain NY12-2, complete genome	CP033576.1	1664	100%	99.56%
	B. velezensis strain DKU_NT_04, complete genome	CP026533.1	1664	100%	99.56%
cce142	*B. amyloliquefaciens* strain KC41, complete genome	CP044444.1	1387	100%	99.48%
	B. velezensis strain NY12-2, complete genome	CP033576.1	1387	100%	99.48%
	B. velezensis strain DKU_NT_04, complete genome	CP026533.1	1387	100%	99.48%
sqo277	*B. amyloliquefaciens* strain KC41, complete genome	CP044444.1	1507	100%	99.88%
	B. velezensis strain NY12-2, complete genome	CP033576.1	1507	100%	99.88%
	B. velezensis strain DKU_NT_04, complete genome	CP026533.1	1507	100%	99.88%
pkl242	*B. amyloliquefaciens* strain KC41, complete genome	CP044444.1	1555	100%	99.19%
	B. velezensis strain NY12-2, complete genome	CP033576.1	1555	100%	99.19%
	B. velezensis strain DKU_NT_04, complete genome	CP026533.1	1555	100%	99.19%
R16	*B. amyloliquefaciens* strain KC41, complete genome	CP044444.1	1535	100%	99.88%
	B. velezensis strain NY12-2, complete genome	CP033576.1	1535	100%	99.88%
	B. velezensis strain DKU_NT_04, complete genome	CP026533.1	1535	100%	99.88%

Table C1. Continued.

Isolate	Top 3 BLAST Matches	GenBank Accession No.	Max / Total Score	Query Cover	Per. Ident
bnd160	*B. velezensis* strain AL7, complete genome	CP045926.1	1698	100%	98.95%
	B. velezensis strain HAB-2, complete genome	CP060085.1	1698	100%	98.95%
	B. velezensis strain ZeaDK315Endobac16, complete genome	CP043809.1	1692	100%	98.85%
ccc103	*B. velezensis* strain AL7, complete genome	CP045926.1	1661	100%	99.78%
	B. velezensis strain HAB-2, complete genome	CP060085.1	1661	100%	99.78%
	B. velezensis strain ZeaDK315Endobac16, complete genome	CP043809.1	1655	100%	99.67%
CL-R51	*B. velezensis* strain AL7, complete genome	CP045926.1	1604	99%	99.44%
	B. velezensis strain HAB-2, complete genome	CP060085.1	1604	99%	99.44%
	B. velezensis strain ZeaDK315Endobac16, complete genome	CP043809.1	1598	99%	99.32%
CP-R23	*B. velezensis* strain AL7, complete genome	CP045926.1	1572	100%	99.31%
	B. velezensis strain HAB-2, complete genome	CP060085.1	1572	100%	99.31%
	B. velezensis strain ZeaDK315Endobac16, complete genome	CP043809.1	1567	100%	99.20%
sqo275	*B. velezensis* strain FJAT-45028, complete genome	CP047157.1	1279	100%	99.86%
	B. velezensis strain WRN014, complete genome	CP041361.1	1279	100%	99.86%
	B. velezensis strain J01, complete genome	CP023133.1	1279	100%	99.86%
bng199	*B. velezensis* strain EN01, complete genome	CP053377.1	1576	100%	99.77%
	B. amyloliquefaciens strain WF02, complete genome	CP053376.1	1576	100%	99.77%
	B. amyloliquefaciens strain DH8030, complete genome	CP041770.1	1576	100%	99.77%
bng230	*B. velezensis* strain OSY-GA1, complete genome	CP031880.1	1474	100%	99.27%
	B. amyloliquefaciens strain S499, complete genome	CP014700.1	1474	100%	99.27%
	B. amyloliquefaciens strain IT-45, complete genome	CP004065.1	1474	100%	99.27%

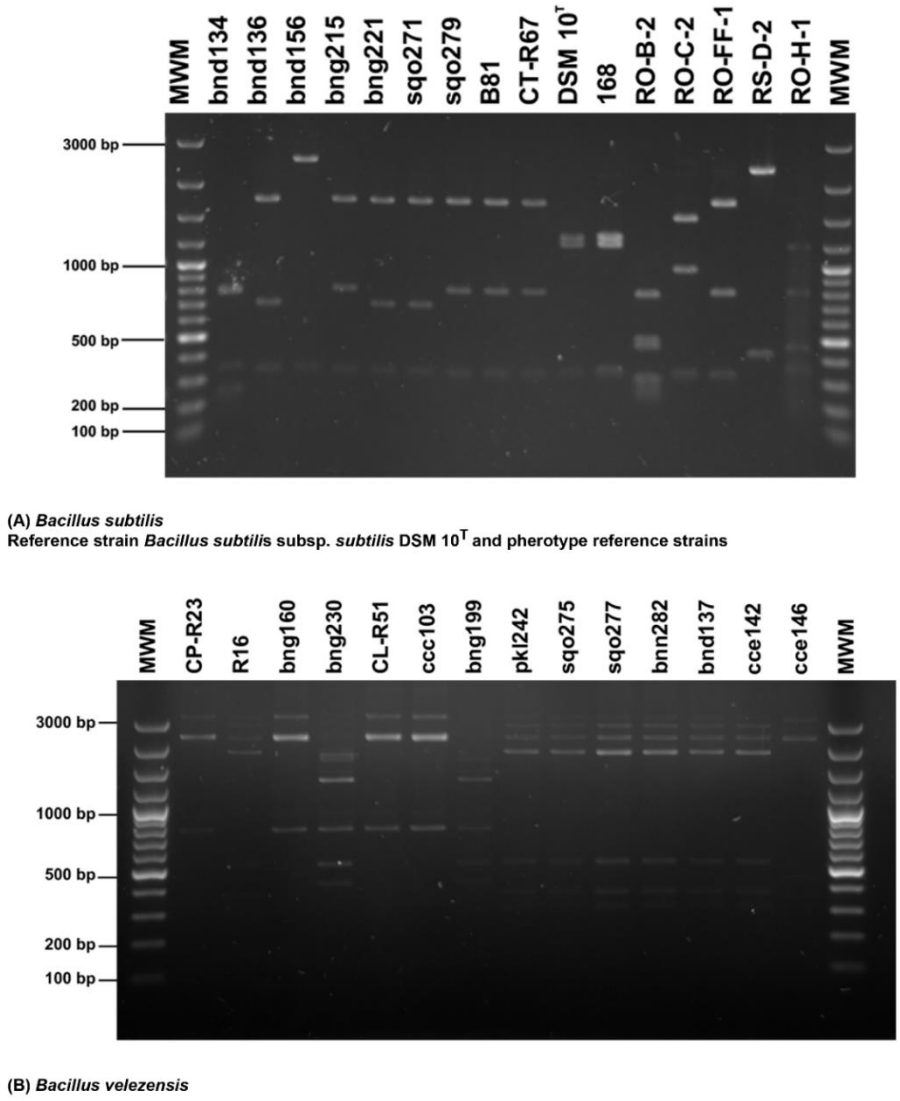
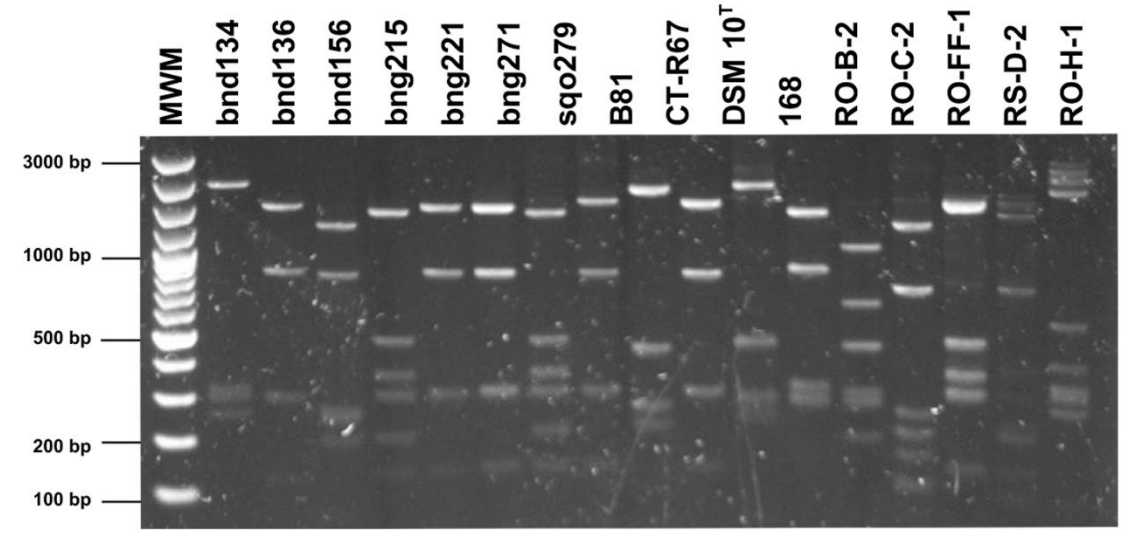
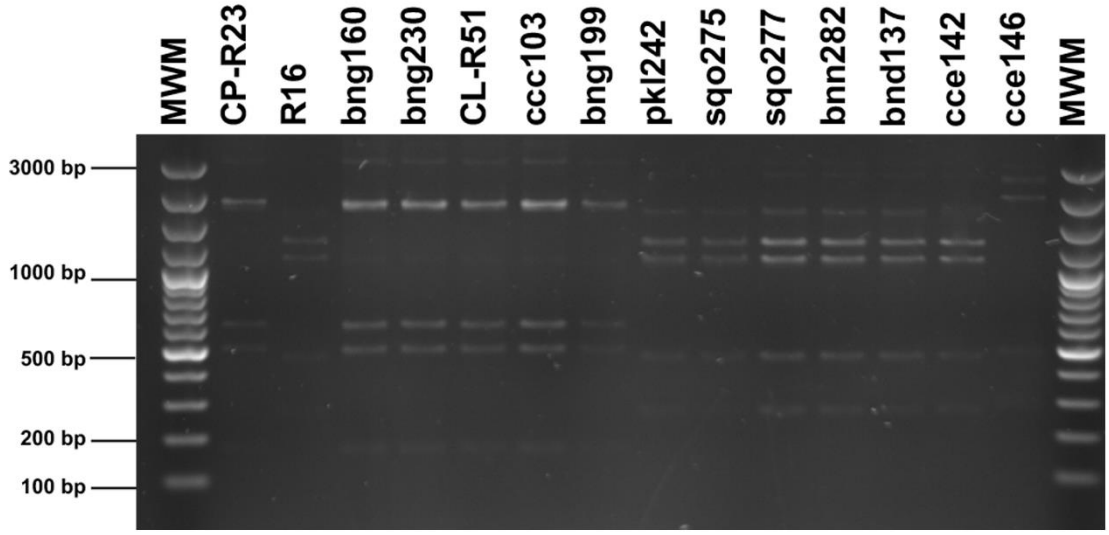


Figure C1. Agarose gel images of *Bts*CI *comQXP* PCR-RFLP profiles of plant-associated *Bacillus subtilis sensu lato* isolates, reference type strains *B. subtilis* subsp. *subtilis* DSM 10^{T}, and pherotype reference strains. Agarose gels at 1 % w/v used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Laboratorois Conda, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 100 bp Plus DNA Ladder (Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".



(A) *Bacillus subtilis*
Reference strain *Bacillus subtilis* subsp. *subtilis* DSM 10^T and pherotype reference strains



(B) *Bacillus velezensis*

Figure C2. Agarose gel images of *Fnu*4HI *comQXP* PCR-RFLP profiles of plant-associated *Bacillus subtilis sensu lato* isolates, reference type strains *B. subtilis* subsp. *subtilis* DSM 10^T, and pherotype reference strains. Agarose gels at 1 % w/v used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Laboratorois Conda, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 100 bp Plus DNA Ladder **(**Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".

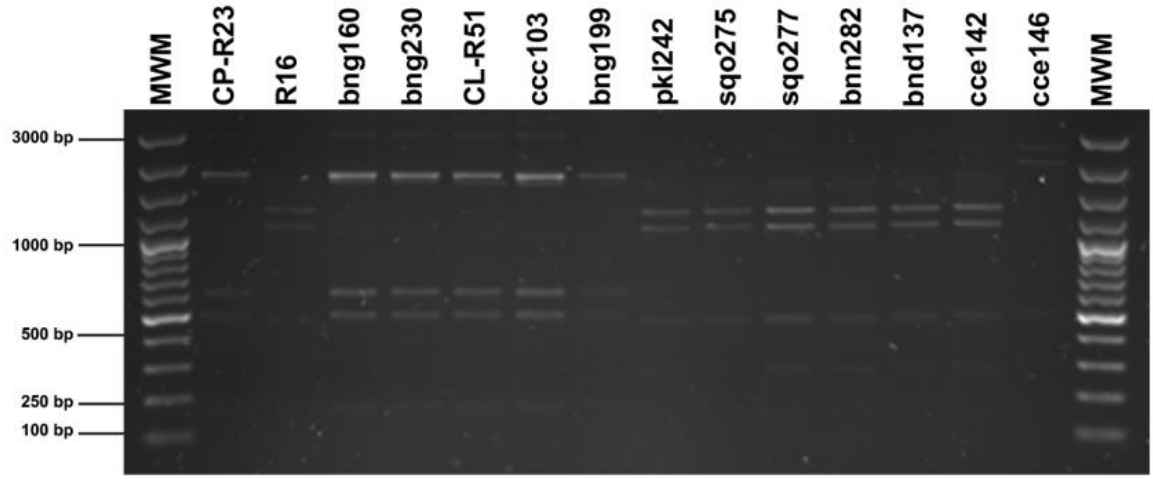


Figure C3. Agarose gel images of *Hpy*166II *comQXP* PCR-RFLP profiles of plant-associated *Bacillus velezensis* isolates. Agarose gels at 1 % w/v used a 1x Tris-Borate-Ethylenediaminetetraacetic acid (TBE) buffer (89 mM Tris base, 89 mM Boric acid and 2 mM EDTA, adjusted to pH 8.0). Agarose gels were pre-stained with 1x CondaSafe (Laboratorois Conda, Madrid, Spain). Amplicon band sizes were determined using a GeneRuler 100 bp Plus DNA Ladder **(**Thermo Fisher Scientific, Inc., Waltham, Massachusetts, U.S.A.) denoted as "MWM".