The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

ON THE PHYLOGENY, DETECTION AND BEHAVIOUR OF CLOSTRIDIAL PATHOGENS IN THE SOIL

JOSEPH SAMUEL PALMER

28th September 2018

Submitted in partial fulfilment for the degree of Ph.D. in Environmental Science at the University of Nottingham

ACKNOWLEDGEMENTS

My sincere gratitude and appreciation goes out to my supervisory team of Dr Lisa Avery, Dr Rupert Hough and Dr Helen West for their academic supervision, but more importantly their regular words of encouragement in the times when I was feeling particularly stupid, and unending patience with me in all the times when I was genuinely being stupid. I am fortunate that my supervisors had faith in me to delve headfirst into the world of pathogens and phylogenetics, facilitating my development as a scientist, and reassured in the knowledge that guidance and superior wisdom was only ever an email away. Lisa and Rupert - thank you for all the editing, for offering me casual work when the stipend disappeared and for the many offers of gardening work to save me from the despairs of my bar job! Helen - how you have put-up with my antics for over 7 years now is mystifying. Thanks for the last-minute edits, the references, off-topic chats and probably most importantly, encouraging me to do this PhD in the first place. I genuinely don't think I would have even contemplated starting (let-alone finishing) my PhD if it wasn't for your belief in me. Therefore, I blame you for these last few months of suffering, but wholeheartedly thank you for facilitating 4 amazing years of development and debauchery.

I would like to thank Dr Adam Wyness, Duncan White and Lucinda Robinson for their amazing support and patience in the laboratory. They trained me, assisted me, and were always happy to help with my (many) questions throughout the years. Additional thanks to Dr Leighton Pritchard and Dr Thomas Freitag for their academic support, and to Emma Hooley for keeping me in contact with base camp south of the border.

Will, Sam, Joe, Ewan and Sophie - I couldn't have asked for kinder, lovelier housemates during my time in Aberdeen, and to Steve H and the motley crew - thanks for all the many drunken shenanigans and for making me feel like a true Aberdonian, "Mon ih Biff!" This PhD would not have been possible without the emotional and financial support of my parents throughout my time in further education, so special thanks to you. Lastly,

the biggest acknowledgment probably needs to go to Maia. You have had the unenviable task of being my partner for the last year and half. This is a relentless task at best of times, but you've really been amazing during these last few months when I have been stressed, tired, and about as charismatic as a garden trowel. I imagine you will be as delighted as anyone when this thesis is done! Truly, from the bottom of my heart, thank you.

CONTENTS

224

LIST OF ABBREVIATIONS

AICc	Akaike information criterion
AFLP	Amplified fragment length polymorphism
AMRISA	Automated multiplex ribosomal intergenic spacer analysis
ANOVA	Analysis of variance
ARISA	Automated ribosomal intergenic spacer analysis
BI	Bayesian Inference
BoNT	Botulinum neurotoxin
BSA	Bovine serum albumin
BTPC	Botulinum-toxin producing Clostridia
CADM	Congruence among distance matrices
CCA	Canonical correspondence analysis
CFU	Colony-forming units
DGGE	Denaturing gradient gel electrophoresis
DM	Distance matrices
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
EDTA	Ethylenediaminetetraacetic acid
EGS	Equine grass sickness
EIA	Enzyme immunoassays
ELISA	Enzyme-linked immunosorbent assays
FA	Fragment analysis
FAM	6-fluorescein
gDNA	Genomic DNA
GLM	Generalised linear model
GLMM	Generalised linear mixed model
GTR	Generalised time-reversible
HK	Housekeeping
IAC	Internal amplification control
ITS	Intergenic transcribed spacer
LH	Length-heterogeneity
LOD	Limit of detection

MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
ML	Maximum-Likelihood
mPCR	Multiplex PCR
N	Nitrogen
NB	Negative binomial
NCS	Nucleotide coding sequences
NMDS	Non-metric multidimensional scaling
NTC	No-template control
OLRE	Observational-level random effects
OSA	Organic soil amendments
PCR	Polymerase chain-reaction
PMA	Propidium monoazide
RA	Relative abundance
rDNA	Ribosomal DNA
RF	Robinson–Foulds
RFU	Relative fluorescence units
RNA	Ribonucleic acid
RT-qPCR	Real-time quantitative polymerase chain-reaction
SASP	small, acid-soluble spore proteins
SDIW	Sterile deionised water
SMB	Standard mouse bioassay
SPR	Subtree prune-and-regraft
SSCP	Single strand conformation polymorphism
TE	Tris-EDTA
TGGE	Temperature gradient gel electrophoresis
TRF	Terminal-restriction fragments
T-RFLP	Terminal-restriction fragment length polymorphism
TYGB	Thioglycollate broth with resazurine
UV	Ultraviolet
VIF	Variance inflation factors
WGS	Whole genome sequencing
WHC	Water holding capacity

ABSTRACT

Clostridial pathogens are the aetiological agents for various human and animal diseases. Their presence in soils, manure and other organic wastes can give rise to multiple routes of infection *via* ingestion, inhalation or contamination of wounds, crops or silage with soil containing vegetative cells or spores. More qualitative and quantitative data are needed on the distribution, abundance and behaviour of these pathogens in agricultural soils to better manage risk.

This project investigated how various soil and land-management factors affect the prevalence and survival of key pathogenic Clostridia in the soil. A multidisciplinary approach was used to develop molecular tools for clostridial diagnostics. For accurate strain identification, single and multi-gene markers were used to produce an updated phylogeny for members of the Class Clostridia, indicating examples of misnamed and misclassified strains. For rapid, simultaneous detection of Clostridia, proof-of-concept for an automated multiplex ribosomal intergenic spacer analysis (AMRISA) protocol was devised, and its utility demonstrated on agricultural soils collected from sites in north-east Scotland. Clinically important Clostridia were identified in soils across farms, including *Paeniclostridium sordellii* and *Clostridium novyi,* which were prevalent in all soils. Farm management type, season and grazing levels affected the clostridial community, with *Pn. sordellii* prevalence and relative abundance highest under dairy and beef regimes. *C. novyi* abundance was highest in the winter. Permanent waterlogging had a pronounced effect on pathogen abundance, although short-term changes in soil moisture had little observable effect. Microcosm studies indicated that *Clostridioides difficile*, *C. perfringens* and *C. sporogenes* were more persistent when spiked into drier soils compared to soils with higher moisture contents. Future work should increase pathogen surveillance on farms and identify the effect of other soil characteristics on pathogen behaviour, to understand disease epidemiology and establish appropriate risk-aversion practices.

CHAPTER 1: GENERAL INTRODUCTION

1.1 HUMAN AND LIVESTOCK PATHOGENS IN THE SOIL

The link between soil and certain diseases is well acknowledged, although not necessarily fully-understood (Oliver and Gregory, 2015). While researchers have been discovering and isolating infectious microbes from soils for well over a century (Noble, 1915; Schoenholz and Meyer, 1922), soil-borne diseases still cause significant loss to life, reductions in health and considerable economic losses globally (WHO, 2016). There are three main pathways through which pathogenic organisms in the soil can directly enter susceptible animals, including humans: ingestion, inhalation or through the skin (including lesions in the skin) (Oliver and Gregory, 2015; Baumgardner, 2012). Pathogens can also be indirectly ingested *via* soil-derived products, such as crops and silage, emphasising the expansive, multifaceted approach that is required for effective disease prevention. While soil-related human disease appears to have been curbed in many developed nations (Jeffery and Van der Putten, 2011), many diseases are a continuing threat in developing countries (Thwaites, Beeching and Newton, 2015; Afshar *et al*., 2011) and for livestock, wildlife and other domestic animals (Songer, 1996; Pirie, Jago and Hudson, 2014; Lewis, 2011; Vidal *et al*., 2013).

One bacterial genus regularly affiliated with soil-borne disease is *Clostridium. Clostridium botulinum, C. perfringens* and *C. tetani* are species frequently reported as major soil pathogens, responsible for particularly debilitating and often-fatal diseases (Hatheway, 1990). Many studies describe these bacteria as euedaphic or 'geo-indigenous' soil pathogens (Jeffery and Van der Putten, 2011; Pepper *et al*., 2009), meaning that they can grow, metabolise and reproduce in the soil. They are also documented as being 'ubiquitous' in soils, implying a uniform, pervasive threat is posed to health from exposure to any soil. However, research into the prevalence of these bacteria in soils, and their adaptation to environmental stressors is limited. To better understand the

epidemiology of clostridial disease, it is imperative to elucidate the behaviour and distribution patterns of these bacteria in the soil.

1.2 CLOSTRIDIAL PATHOGENS

Clostridium is a large genus, typically characterised by endospore-forming, rod-shaped, Gram-positive and anaerobic bacteria. The genus contains over 180 known species, exhibiting considerable phenotypic and genetic diversity (Hatheway, 1990; Bahl and Dürre, 2001; Dürre, 2004; Keto-Timonen *et al*., 2006). *Clostridium* belongs to the Clostridia class of bacteria within the Firmicutes phylum. Although many *Clostridium* spp. are pathogenically benign, at least 10 species are known to induce disease in humans or other animals (Figure 1.1) (Popoff and Bouvet, 2013; Collins *et al*., 1994; Stackebrandt *et al*., 1999). *Clostridium* pathogens contribute significantly to the global burden of disease, in part, due to the ubiquity of the organism in many environments, the potent toxins produced and the longevity of the environmentally-persistent endospores (Hatheway, 1990). For example, despite highly effective vaccination programs, tetanus (caused by *C. tetani*) was responsible for over 10,000 confirmed deaths in 2015 alone (WHO, 2016). *C. perfringens* is a leading cause of gastroenteritis, with 16 reported outbreaks and 510 individual cases diagnosed in England and Wales during 2013 (Hughes, Gillespie and O'Brien, 2007; England, 2016). Other significant *Clostridium* diseases include botulism (*C. botulinum, C. baratii* and *C. butyricum*) (Espelund and Klaveness, 2014; Fach *et al*., 2011) and various gangrenous and necrotic conditions (*C. perfringens, C. novyi, C. septicum* and *C. chauvoei*) (Sasaki *et al.*, 2000; Brynestad and Granum, 2002; Lindström *et al.*, 2011; Skarin and Segerman, 2014). The formerly-assigned *Clostridium* spp. *Clostridioides difficile, Paeniclostridium sordellii* and *Paraclostridium bifermentans* are genetically and phenotypically similar to *Clostridium sensu stricto* species, and due to their significant role in disease mediation, they will also be discussed within this review. From herein, the term 'clostridial pathogens' is used to describe all key *Clostridium sensu stricto*

pathogens, as well as the closely-related infectious *Cl. difficile*, *Pn. sordellii,* and *Pr. bifermentans* species. Clostridial bacteria are also the aetiological agents for important veterinary diseases including; blackleg in cattle, sheep and swine (*C. chauvoei*) (Sasaki *et al*., 2002), enterocolitis in horses, lamb dysentery and necrotic enteritis in piglets and poultry, enterotoxaemia in sheep, goats and foals (*C. perfringens*) (Van Immerseel *et al*., 2004), equine grass sickness (EGS) (*C. botulinum*) and various other diseases (Hunter, Miller and Poxton, 1999; Mccarthy *et al*., 2010). These diseases can incur substantial financial losses for the animal owners and may contribute to an overall reduction in animal welfare (Bagge, Persson and Johansson, 2010).

1.3 PATHOGEN RESERVOIRS AND ABUNDANCE

Pathogenic Clostridia are often carried commensally/asymptomatically as part of the indigenous microflora of the mammalian gut, and are frequently described as ubiquitous in soil, sediments and decaying organic matter (Matches, Liston and Curran, 1974; Hatheway, 1990; Hielm *et al*., 1996). In the soil, Clostridia are part of a microbial consortium that play a pivotal role in nutrient recycling of organic and inorganic compounds, improving soil fertility, boosting plant health and other soil functions (Garbeva, van Veen and van Elsas, 2004; Kirk *et al*., 2004; Ulrich and Becker, 2006). Soil is thought to be the major reservoir for many pathogenic strains and species (Meng *et al*., 1999; Li, Sayeed and McClane, 2007; Mccarthy *et al*., 2010), representing an important pathway in disease transmission to human food products and grazing animals. However, there is only limited research on the prevalence and diversity of clostridial species in the soil (Stefanis *et al*., 2014). Previous research has focused on determining the prevalence of individual, key pathogenic species or toxinotypes in often spatially-limited studies. A summary of pertinent clostridial soil studies is shown in Table 1.1. Other epidemiological studies have focused on linking the prevalence of specific clostridial species in the soil to

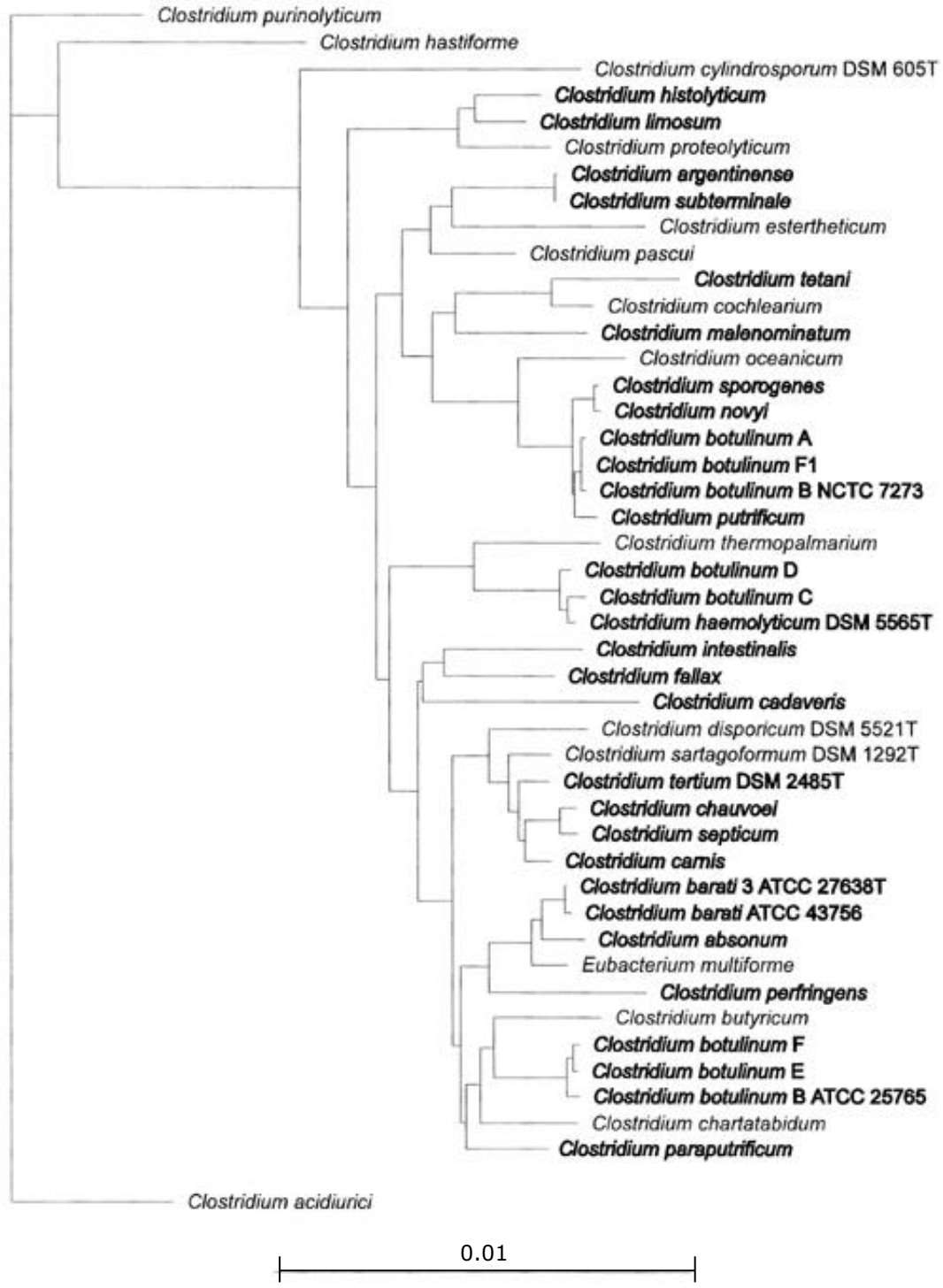


Figure 1.1 Position of suspected pathogenic *Clostridium sensu stricto* species and strains (in bold) next to their non-pathogenic neighbours within 16S rDNA cluster I. Scale bar represents 10% sequence divergence (Stackebrandt *et al*., 1999).

the incidence of certain diseases, normally in case-control studies. Despite the large body of research in clostridial pathogen presence, the studies have failed to identify the main soil determinants affecting the presence and abundance of these pathogens. Furthermore, these studies utilised a range of different microbiological techniques, which may limit the scope for integrating results, for example by using meta-analysis techniques.

1.4 CLOSTRIDIAL TAXONOMY AND IDENTIFICATION

Taxonomy and detection of *Clostridium* species based on phenotypic traits has often proved inaccurate or misleading due to high phenotypic heterogeneity within these features (Bahl and Dürre, 2001; Collins *et al*., 1994). However, the increasing availability of genomic nucleotide sequences and advances in cluster analysis approaches offer an increasingly comprehensive methodology for taxonomic reorganisation of the genus (Collins *et al*., 1994; Yutin and Galperin, 2013). Collins *et al*. (1994) used ribosomal 16S gene sequences to explore *Clostridium* phylogeny, leading to a major revision of the *Clostridium* genus. Species were tentatively assigned to 19 distinct family-level groups (Clusters I - XIX) using distance matrixes, neighbour-joining and bootstrapping analytical techniques (Harrison and Langdale, 2006). These clusters represent phylogenetically distinct families, although the *Clostridium* genus also demonstrates considerable phenotypic diversity. For example, Gram-negative, non-spore-forming and non-anaerobic species are now included in the genus, which defies many historical characteristics of *Clostridium* (Finegold, Song and Liu, 2002; Yutin and Galperin, 2013). The genetic relatedness of some *Clostridium sensu stricto* (Cluster I) species is shown in Figure 1.1. Most of the significant pathogens covered in this review are found in this cluster. Notable exceptions are *Cl. difficile, Pn. sordellii* and *Pr. bifermentans,* which have recently been reassigned to novel genera within family *Peptostreptococcaceae* (along with other Cluster XI species),

to better represent the evolutionary relationships of these pathogens (Yutin and Galperin, 2013; Lawson *et al*., 2016).

1.5 STRAINS AND TOXINOTYPES

Many species of Clostridia can be further differentiated into strains and toxinotypes depending on growth preferences or the presence of toxin-encoding genes. Clostridial diseases are mediated by the production of extracellular toxins. For example, eight botulinum neurotoxin (BoNT) serotypes (Types A-H) have been currently identified (Peck *et al*., 2017), causing various forms of botulism. BoNT Type A is the most potent toxin in existence (Brunt *et al*., 2014; Hatheway, 1990). The presence of toxins, or toxin-coding genes has often been used to distinguish between species and strains, such as for the differentiation between *C. botulinum* and the non-pathogenic *C. sporogenes* (Weigand *et al.*, 2015). Different toxin combinations can induce specific syndromes (Petit, Gibert and Popoff, 1999), or affect the virulence of strains (Persson, Torpdahl and Olsen, 2008). In *Cl. difficile,* nucleotide variations occur between strains in a genomic region called the pathogenicity locus (PaLoc) which effects the expression of toxin genes *tcdA* and *tcdB*, and a toxin-regulator gene *tcdC,* causing significant variations in strain virulence (Popoff and Bouvet, 2013; Dingle *et al*., 2011; Griffiths *et al*., 2010). Many strains are classified by toxinotype, based on the major toxins produced, such as with *C. botulinum, C. perfringens, C. novyi* and *Cl. difficile* (Hielm *et al*., 1996; Brynestad and Granum, 2002; Popoff and Bouvet, 2013; Hatheway, 1990). Some toxins can be produced by multiple *Clostridium* species. *C. baratii* can produce BoNT Type F like Group I and II *C. botulinum,* while *C. butyricum* can produce BoNT Type E (Popoff and Bouvet, 2013; Smith *et al*., 2015). Species that produce BoNTs are collectively termed botulinum-toxin producing Clostridia (BTPC) (Lúquez *et al*., 2005). Some toxin genes are carried by bacteriophages, such as the BoNT Type C, D and C/D or D/C chimeric variant genes (Moriishi *et al*., 1996; Skarin and Segerman, 2014), and toxin genes can be located within chromosomal or plasmid DNA.

Table 1.1 Summary of pertinent studies of Clostridia soil prevalence

Species	Author	Sample area(s)	No. of samples
C. argentinense	Sonnabend *et al*. (1987)	Switzerland	41
C. botulinum	Creti *et al*. (1990)	Rome	520
	Huss (1980)	Denmark, Faroe Is. & Iceland	118
	Lúquez *et al.* (2005)	Argentina	2009
	Serikawa *et al*. (1977)	Japan	230
	Smith and Milligan (1979)	London	60
	Smith and Young (1980)	Great Britain	74
	Yamakawa *et al*. (1988)	Japan & Shinkiang Province (China)	286
C. butyricium	Meng *et al*. (1999)	Jiangsu Province (China)	60
Cl. difficile	Al Saif and Brazier (1996)	South Wales (UK)	104
	Båverud *et al*. (2010)	Sweden	598
C. perfringens	Kuske *et al*. (2006)	USA	129
	Li *et al.* (2007)	Pittsburgh (USA)	502
	Voidarou *et al*. (2011)	Greece	750
C. tetani	Smith (1978)	USA	260
	Wilkins *et al*. (1988)	South Africa	60
Multiple	Gamboa *et al.* (2005)	Costa Rica	117
	Hang'ombe *et al.* (2000)	Zambia	46
	Miwa (1975)	Antarctica	31
	Sathish and Swaminathan (2009)	Southern India	115

Recent studies have demonstrated the ability of toxin genes to be transferred to genetically related, non-toxigenic species, and *vice versa* (Skarin and Segerman, 2014; Weigand *et al*., 2015). Studies have observed different growth preferences and geographical abundances between clinically distinct strains and toxinotypes. This highlights the complex but poorly-understood relationships between the strain physiology and genotype, growth and survival in soil, and disease acquisition (Hatheway, 1990). Determining the main drivers behind species and strain growth in different soils is essential for understanding clostridial disease epidemiology.

1.6 PROJECT AIMS AND OBJECTIVES

The aims of this thesis were to elucidate the underlying distribution, abundance and behaviour of pathogenic Clostridia in soils, with an emphasis on agricultural soils, where the risk of disease transmission is likely to be enhanced.

To achieve these aims, the following objectives were identified:

1. Review existing literature on clostridial epidemiology and prevalence in the soil to establish potential links between disease, pathogen behaviour and soil factors.
2. Optimise or design appropriate methodologies for the accurate identification and detection of clinically important Clostridia.
3. Determine the influence of soil characteristics on the growth of key clostridial pathogens.
4. Assess the distribution and diversity of clostridial pathogens in agricultural soils.
5. Explore the impacts of farming regime on clostridial pathogens in the soil.

1.7 THESIS STRUCTURE

The thesis was written by the first author, Joseph Samuel Palmer, and edited by supervisor's Dr Lisa Avery, Dr Rupert Hough and Dr Helen West. All experimental work was performed by Joseph Palmer at the James Hutton Institute laboratories in Aberdeen. Fieldwork was conducted by Joseph Palmer at James Hutton Institute estates or other agricultural sites, with the express permission of landowners. The thesis addresses the aforementioned aims in the following chapters, concluding with a general discussion (Chapter in which knowledge outputs and future research ideas are expanded:

Chapter 2: Prevalence and distribution of pathogenic Clostridia in soil: A Review

Research pertaining to the prevalence and behaviour of clostridial pathogens in the soil was reviewed, including the identification of risk factors associated with clostridial disease in horses and livestock. The biotic and abiotic factors likely to affect clostridial growth in soil were identified in the review. The chapter concludes with an evaluation of the diagnostic techniques for the detection and study of clinically important Clostridia and recommendations for future method standardisation and harmonisation are made.

Chapter 3: A review of clostridial phylogeny using multilocus sequence analysis

Nomenclatural conflicts within the class Clostridia can be prohibitive for effective research in this field. To this end, multilocus phylogeny reconstruction of 101 fully sequenced clostridial genomes was performed to identify other potential taxonomic inaccuracies. A further aim was to identify suitable-single markers that would allow rapid, accurate taxonomic assignment, with the multilocus phylogeny used as a benchmark.

Chapter 4: Survival of pathogenic Clostridia in water amended soil

The aim of this study was to ascertain whether waterlogged soils could sustain a high pathogen abundance post-spiking. Real-time quantitative PCR was used to monitor the survival of *Cl. difficile*, *C. perfringens* and *C. sporogenes* in agricultural soils over an 84-day period and from this, pathogen die-off coefficients were derived.

Chapter 5: Development and first evaluation of novel AMRISA protocol for rapid, simultaneous detection of clostridial pathogens

A proof-of-concept study was used to design and test PCR-based molecular diagnostics that could facilitate the sensitive, simultaneous detection of key clostridial pathogens. Candidate assays were thoroughly tested *in silico*, before assessing the limits of detection, sensitivity and specificity of the tests *in vitro*.

Chapter 6: Characterisation of the presence and behaviour of Clostridia in Scottish agricultural soils

A field sampling campaign was used to explore the temporal and spatial changes in the prevalence and relative abundance of clostridial pathogens in Scottish agricultural soils. The study aimed to explore the prevalence and behaviour of these bacteria across agricultural sites under different management practices. Seasonal changes could be assessed by the 6-week sampling regime, while the affect of persistent waterlogging was determined by the inclusion of waterlogged/well-drained sampling sites. Pathogen detection and identification was achieved using custom-designed molecular diagnostics designed in the previous chapter.

CHAPTER 2: PREVALENCE AND DISTRIBUTION OF PATHOGENIC CLOSTRIDIA IN SOIL: A REVIEW

2.1 PATHOGEN DIVERSITY AND PREVALENCE

Clostridia have often been considered a common constituent of soil microflora (Russo *et al*., 2012; Janssen, 2006). In a survey of 16S rDNA gene libraries, on average, 0.59% of the soil bacteria community belonged to Clostridia in the 3,398 gene clones examined in 32 gene libraries (Janssen, 2006). However, the species, strain and toxinotype diversity of Clostridia can vary significantly between locations; from large-scale continental differences (Dodds, 1992; Haagsma, 1991) and regional differences (Smith, 1978; Smith and Young, 1980; Lúquez *et al*., 2005; del Mar Gamboa, Rodríguez and Vargas, 2005), through to microscale differences within soil of the same sample (Kirk *et al*., 2004).

Gamboa *et al*. (2005) studied the prevalence of Clostridia in Costa Rican soils, revealing 54 different clostridial species in 117 samples, averaging over 7 species per sample. The most prevalent species were *C. subterminale, C. oceanicum* and *C. glycolicum.* Additionally, 11 toxigenic species were isolated, with *Pn. sordellii* and *C. perfringens* being the most prevalent pathogens (present in 42% and 38% of samples, respectively). *C. tetani* (4% of samples), *Cl. difficile* (3% of samples) and *C. botulinum* (1% of samples) were also identified. Kim *et al*. (2004) isolated 16 different *Clostridium* species from 152 South Korean soil samples across 5 locations. *C. perfringens* was common across all sampling locations, indicating the ubiquity of this pathogen across different soil types and agricultural regimes. Other pathogens, such as *C. chauvoei*, *C. novyi, C. septicum* and *Cl. difficile* were only detected in certain locations. However, *C. botulinum* was not isolated from any soil, possibly due to soil alkalinity and other environmental factors (Kim *et al*., 2004).

2.2 STRAIN, TOXINOTYPE AND CELLULAR FORM VARIATIONS

C. botulinum prevalence has been examined more widely, particularly in American and European soils. Some *botulinum* producing toxinotypes (BTPC) appear endemic to certain geographical areas or environments. Based on the literature, toxin Type A is frequently isolated in North American soils west of the Mississippi river and in uncultivated soils, whereas Type B prevails in soils to the east of the Mississippi, European soils and cultivated soils (Dodds, 1992; Espelund and Klaveness, 2014; Haagsma, 1991). Types C, D, G and C/D mosaic strains are also common in European soils, with C, D and C/D mosaic strains frequently associated with environmental botulism outbreaks. Type E is commonly isolated from marine environments, such as fish gut contents and coastal sediments, with research indicating this toxinotype has a higher affinity for permanently wet environments (Haagsma, 1991; Espelund and Klaveness, 2014). Lúquez *et al*. (2005) surveyed the prevalence of BTPC in 2,009 Argentinean soil samples. BTPC were identified in 23.5% of samples, which is comparatively high compared to studies in other locations, such as the 5.7% and 16.5% prevalence found in British (Smith and Young, 1980) and Japanese soils respectively (Yamakawa *et al*., 1988). *C. perfringens* spores were between 30-65% prevalent in Greek soil samples depending on the type of overlying arable cultivation (Voidarou *et al*., 2011), with the spore form (47.5%) more prevalent than the vegetative form (11%). A study of Greek soils found *C. perfringens* spores and vegetative cells in 36.4% and 25.5% of soils, respectively (n =110) (Stefanis *et al*., 2014). *Cl. difficile* was isolated in 21% of Welsh soil samples (Al Saif and Brazier, 1996), and in 37% of 147 soils samples from a rural Zimbabwean homestead (Simango, 2006). *C. tetani* prevalence ranged from 25-42% in 5 different world-wide studies (total of 2,491 soil samples) (Wilkins *et al*., 1988). These studies demonstrate how prevalent many clostridial pathogens are in various intensive and extensive agricultural environments, and why a

better understanding of clostridial behaviour in soils would be of global impact.

2.3 CLOSTRIDIA IN AGRICULTURAL SOILS

Agricultural soils are a reservoir for many food-borne pathogens (Newell *et al*., 2010), including clostridial pathogens, and represent the first critical control point in the food-contamination pathway (Stefanis *et al*., 2014). Pathogens may be acquired by humans and susceptible animals from the soil or vegetation, through wound infection, ingestion or inhalation (Haagsma, 1991; Baumgardner, 2012). Produce can be contaminated by both vegetative cells and spores, both of which can induce disease (Tabaqchali and Jumaa, 1995). Moreover, some studies have found that clostridial species constitute part of the endophytic plant population, suggesting Clostridia may demonstrate mechanisms of internalisation within plant tissue (Miyamoto, Kawahara and Minamisawa, 2004). The longevity of clostridial spores in the soil increases the time-window for bacterial transmission (Tabaqchali and Jumaa, 1995; Girardin *et al*., 2005; Gessler and Böhnel, 2006). Girardin *et al.* (2005) demonstrated the longevity of *C. sporogenes* spores in the soil after application (>1 year), and their subsequent transfer to parsley plants growing in the soil. Due to the substantial virulence of many species, and the high likelihood of many of these pathogens being in agricultural soils, a better understanding of Clostridia survival and behaviour in agricultural environments is imperative to reduce the risk of food borne disease and grazing-livestock illness. Additionally, discrimination between spore and vegetative forms will allow the development of more accurate and effective microbial risk assessment tools.

2.4 FACTORS AFFECTING THE SURVIVAL AND GROWTH OF CLOSTRIDIA

Investigations into soil microbial diversity using molecular techniques have indicated that the soil type, environmental factors, and agricultural management of a soil can influence the microbial community, including the growth and persistence of pathogenic microbes (Roesch *et al*., 2007; Baumgardner, 2012). The five major environmental factors expected to influence the abundance, behaviour and die-off of bacteria in soil are temperature, moisture content, UV exposure, oxygen concentration and biotic interactions (Venglovsky, Martinez and Placha, 2006). This section is an overview of relevant studies that have investigated the effect of these factors on pathogen behaviour, or specifically on clostridial pathogens where available.

2.4.1 Temperature

Clostridia demonstrate a wide range of temperature optima, with psychrophilic, mesophilic and thermophilic representatives. The majority, including most key pathogenic species, grow optimally between 30 and 40°C, optima suited to the temperatures of animal hosts where they are often carried enterically (Bahl and Dürre, 2001; Minton and Clarke, 1989). Additionally, it has been shown that the optimum conditions for growth, sporulation and germination can vary between strains of the same species (Jensen, Genigeorgis and Lindroth, 1987). The culmination of these effects make behaviour and population dynamics in the soil difficult to predict (Evans *et al*., 1997). Most *C. perfringens* strains have a generation time of less than 20 minutes at temperatures between 33 and 49°C, although the pathogen is capable of growing between 15–55°C (Brynestad and Granum, 2002; Albrecht, 2005). Grecz and Arvay (1982) found the optimum temperature for *C. botulinum* (Type E) was 32.5°C, although detectable growth of vegetative cells was observed between 6–41°C. Physiological differences between the *C. botulinum* groups allow differing growth temperature optima, which may explain the different prevalence

and composition of *C. botulinum* groups in different climatic areas. The onset of equine grass sickness (EGS) (thought to be caused by *C. botulinum* Type C*)* in the UK has a strong seasonal occurrence. The temperature optima for pathogenic *C. botulinum* groups and other important clostridial pathogens are summarised in Table 2.1. Interestingly, Miwa (1975) identified *C. butyricum, C. perfringens*, *C. septicum* and *Pn. sordellii* pathogens in Antarctic soils exposed to temperatures between -38°C and +3.2°C. This not only indicates the presence of pathogens in extreme environmental conditions, but also demonstrates their ability to survive outside laboratory-determined temperature ranges.

There are only limited studies examining clostridial growth in soil, and the interaction between temperature and soil properties. Studies indicate that *C. perfringens* abundance varies seasonally, with population decreases observed after cold periods (Brochier *et al*., 2012). However, *C. perfringens* spores remain viable in agricultural soils despite exposure to sub-optimal growth conditions (Garrec, Picard-Bonnaud and Pourcher, 2003). Brocklehurst and Lund (1982) isolated 13 strains of pectinolytic Clostridia from UK soils, 6 of which could grow at < 5 °C. Similarly, Perry (1985) found clostridial populations of 4.4–23.5 x 10^3 CFU g^{-1} from Scottish soils between October-November. Most isolates from the soils (16/24) were capable of significant growth at 10°C. *C. botulinum* abundance is also likely to show strong seasonal fluctuations, a phenomenon supported by the strong seasonality in botulism-borne diseases such as avian botulism outbreaks and EGS (Sandler *et al*., 1993; Mccarthy *et al*., 2010; Espelund and Klaveness, 2014). Incidence of EGS peaks in the springtime, particularly in May (Wylie and Proudman, 2009; Wood, Milne and Doxey, 1998), suggesting that warm conditions may be a predisposing factor for *C. botulinum* proliferation in soil.

Table 2.1 Summary of *in vitro* optimum pH and temperature growth conditions for selected *Clostridium* pathogens

Organism	Temperature optima °C	Range °C	Optimum pH	Reference
C. botulinum				
Group I (proteolytic)	35–40	10–48	4.6–9	McLauchlin and Grant (2007); Stringer *et al*. (2013)
Group II (nonproteolytic)	18–30	2.5–45	5.0–9	McLauchlin and Grant (2007); Stringer *et al.* (2013)
C. butyricum	n/a	8	4.22*	Ghoddusi, Sherburn and Aboaba (2013)
C. histolyticum	30–37	< 25–45	8.5	Whitman and Parte (2009)
C. novyi	45	>25	< 8.5	Whitman and Parte (2009)
C. perfringens	43–47	15–55	5–9	Albrecht (2005)
C. tetani	37	14–43	7.4	Chessbrough (2002)
Cl. difficile	30–37	> 25 – <45	6·5–7·5†	Wheeldon *et al.* (2008); Whitman and Parte (2009)
Pn. sordellii	37†	> 25 – <40 †	5.7–6.5†	Ramirez and Abel-Santos (2010)

* Minimum pH for growth

† Optimal conditions for spore germination

2.4.2 Moisture

Soil moisture content is mainly governed by local precipitation patterns or irrigation methods. The soil moisture content will also influence the soil oxygen-status. Saturated soils will give rise to oxygen-depleted (anoxic) environments, favouring the growth of anaerobes. Båverud *et al.* (2010) identified a higher frequency of *Cl. difficile* in soil samples from water-filled ditches. This phenomenon is supported by numerous studies into bacterial

populations in saturated (i.e. anaerobic) rice paddy soils. Weber *et al.* (2001) attributed 20 out of 31 clones isolated from soil using denaturing gradient gel electrophoresis (DGGE) to the Clostridia class. Additionally, they found 55% of active cells detected after 8 days anaerobic incubation of paddy soils belonged to the *Clostridium* genus. A review of flooded rice paddy microbiology by Liesack *et al.* (2000) also concluded that Clostridia and Clostridia-like lineages of bacteria are typical inhabitants of flooded paddy soils. The prevalence of pathogenic species in sporadically-waterlogged pasture and arable (non-paddy) soils is an important yet under-researched line of enquiry. Pathogenic strains are prevalent in many anoxic environments such as marshes, mudflats and water-bodies (Sandler *et al*., 1993; Smith, Milligan and Moryson, 1978). *C. botulinum* Type C, the aetiological agent of avian botulism epizootics, was isolated in over half of 2,200 sediment samples from 10 marshes in a study by Sandler *et al.* (1993). They also found a higher prevalence of the pathogen in permanently vs. seasonally flooded marshes. Mccarthy *et al.* (2010) suggest a decreased risk of EGS in pastures where soil drainage is utilised, supporting the association between higher clostridial disease risk and wet or waterlogged soils. Localised areas of perpetually waterlogged soil could act as contamination 'hotspots' in both arable and livestock regimes. Highly-poached, low-lying areas in grazed fields, such as around feeders or drinking troughs could be at particularly high-risk of contamination.

2.4.3 Light Exposure

On the soil surface, Clostridia and other potential soil pathogens will be exposed to varying amounts of UV radiation from sunlight. At certain wavelengths, bacteria are inactivated due to DNA/RNA damage which prevents cellular processes such as translation and transcription, inhibiting multiplication (Hijnen, Beerendonk and Medema, 2006; Gehr *et al*., 2003). UV radiation is used to disinfect potable water, municipal wastewater, wastewater and greywater effluent (Gilboa and Friedler, 2008; Hijnen, Beerendonk and Medema, 2006). However, *C. perfringens* demonstrates greater UV light resistance than other pathogenic indicators (Gehr *et al*.,

2003), and spores are more resistant than vegetative cells (Hijnen, Beerendonk and Medema, 2006). Spore resistance to UV light (in addition to other environmental factors) is mainly due to the high concentration (5–10%) of α/β-type small, acid-soluble spore proteins (SASP) in the spore core, which protect the DNA backbone from damage (Setlow, 2007). Lanao *et al.* (2010) observed a 1.2 log inactivation of *C. perfringens* vegetative cells in river water, after a 5-minute exposure to light (λ 320–800 nm), whereas spores were only inactivated by <0.5 log after 30 minutes light exposure. Importantly, the light penetration depth (depth at which light intensity at surface is reduced by 99%) in soils can be as little as 300 μm (Ciani, Goss and Schwarzenbach, 2005). Therefore, solar inactivation may represent a small but effective method of vegetative cell destruction, but only at the soil surface. Tillage could enhance clostridial die-off by exposing deeper soil layers to UV radiation.

2.4.4 Oxygen availability

Clostridia are frequently described as obligatory anaerobic organisms, although many species can exhibit varying degrees of aerotolerance (Brioukhanov and Netrusov, 2007). Studies have shown that pathogenic species including *C. botulinum* (Dezfulian, 1999), *C. perfringens* (Briolat and Reysset, 2002) and *C. tetani* (Brüggemann *et al*., 2004) can survive in temporarily microoxic environments in a growth-arrested stage. This adaptive response to oxidative stress is thought to be dependent on a range of specialised genes, some which are permanently expressed, such as rubrerythrins (*rbr*) and superoxide dismutases (*sod*), or some which are transcribed under oxidative stress, including alkyl hydroperoxide reductases (*ahpC*) and heme oygenases (*HemT*) (Briolat and Reysset, 2002; Brüggemann *et al*., 2004; Jean, Briolat and Reysset, 2004; Brioukhanov and Netrusov, 2007). The resultant effect of these defensive mechanisms is to increase clostridial resistance to oxygen-exposure whilst in a vegetative form. However, long-term exposure to oxygen will cause cell mortality. *In vitro* studies have shown that 18 hours of oxygen exposure (100% O_2 at 3 atmospheres absolute pressure) results in almost

complete inactivation of *C. perfringens, C. histolyticum, C. novyi* and *C. tetani* vegetative cells, although *Pr. bifermentans*, *C. butyricum* and *C. septicum* were slightly more aerotolerant (Hill and Osterhout, 1972; Brioukhanov and Netrusov, 2007). Variable oxygen-concentrations in soil micropores may provide niches for the vegetative cells to survive within most soils. However, clostridial spores are highly resistant to oxygenic species, which is again mainly due to high α/β-type SASP concentrations in the spore core (Setlow, 2007). SASP protect DNA against damage from hydroxyl radicals and hydrogen peroxide originating from atmospheric oxygen.

2.4.5 Biotic interactions

Another important consideration is the effect of biological interactions on the survival and behaviour of pathogens in the soil. This is arguably the most difficult factor to interpret, predict or model *in situ* due to the various combinations of mutualistic, neutralistic and antagonistic ecological interactions, and the vast number of other possible indigenous and/or introduced soil microorganisms (Moynihan, 2012). One such interaction is the predation of pathogens by bacteriophages (phages), which are likely to be present in the same natural habitats as the bacteria (Ogata and Hongo, 1980; Minton and Clarke, 1989). Numerous phages have been identified for most pathogenic *Clostridium* species, and they are likely to play an important role in population control. A study of prophage (phage DNA integrated into bacterium DNA) carriage in estuarine sediment revealed that 74% of *Cl. difficile* carried phage particles, demonstrating the key role that phages play in the bacterial life cycle (Hargreaves *et al*., 2013). On the contrary, an earlier study failed to identify any *Cl. difficile* phages in soil, animal faeces or sewage samples (Goh, Riley and Chang, 2005). Goh *et al.* (2005) attributed this finding to the fact that the *Cl. difficile* is often found in spore form in environmental samples, and phages require the host bacterium to be in an active (vegetative) stage of growth for phage multiplication. Additionally, spores lack cell surface structures such as pili required for phage-reception, which may infer a higher

resistance of spores to phage infection. It has also been demonstrated that phages can influence the toxin production of some pathogenic *Clostridium* (Minton and Clarke, 1989).

Competitive inhibition by other microorganisms is also an important factor for clostridial die-off. Soil microbes have been shown to produce antimicrobial products, such as bacteriocins (ribosomally synthesised antagonistic peptides produced by bacteria), which kill or inhibit growth of other bacteria (He, Chen and Liu, 2006). These antimicrobial compounds can show either inter or intraspecific inhibition. Potential soil bacteria such as *Enterococcus fecalis*, *E. faecium*, *Bacillus badius, B. mycoides, B. cereus* and several *Streptococcus* species have been shown to inhibit growth and toxin production of various *C. botulinum* strains (Smith, 1975b; Sandler, Rocke and Yuill, 1998; Shehata *et al.*, 2013). Smith (1975a) also demonstrated the inhibitory effect of *C. perfringens* isolated from the soil on *C. botulinum* and *C. sporogenes* growth. A study of 10 Californian marshes revealed that 32% of samples (n=1600) contained bacteria inhibitory to *C. botulinum* Type C, although there was no statistical relationship between the prevalence of *C. botulinum* and the presence of growth-inhibiting bacteria (Sandler, Rocke and Yuill, 1998). These observations could suggest that only some pathogens are prevalent in certain soils, as opposed to the prevalence of multiple species. *C. botulinum* may be reduced or absent in soils with *C. perfringens* populations. Conversely, research by Sandler *et al.* (1998) suggests that the presence of antagonistic bacteria has a negligible role in the prevalence of *C. botulinum*. Further research is needed to identify other antagonistic relationships between *Clostridium* pathogens and other bacteria, and their significance in controlling pathogen prevalence and abundance.

2.5 SOIL PHYSICOCHEMICAL FACTORS

2.5.1 Soil type and structure

Soil type and structure are two factors which have a strong effect on the prevalence, survival and movement of bacteria in all soils. The soil parent material will influence various soil properties including soil chemistry, texture, porosity and nutrient availability. These features are, in part, controlled by variations in the composition of the main mineral components; clay, silt, sand and organic matter. Some studies indicate that clay and silt-dominated soils allow a greater bacterial diversity and biomass (Russo *et al*., 2012; Ulrich and Becker, 2006). However, this does not necessarily infer a higher prevalence of clostridial species in these soils. In a review by Mawdsley *et al.* (1995) of pathogens in the soil, the authors state that as bacterial populations are confined to the aqueous phase and solid-liquid interface of soils, soil water content and water movement are of upmost importance to bacterial survival and movement. Organic soils have been shown to extend the survival of faecal coliforms in soil, which is likely due to the enhanced water-holding capacity of these soils (Unc and Goss, 2004). Similarly, the composition of other colloidal materials such as aluminium oxides and clay particles facilitate adsorption of cations and water molecules, creating microenvironments more favourable for growth. Bacterial retention by direct cell adsorption to charged soil particles also occurs, with adsorption rates influenced by a multiplicity of factors including the amount and type of colloidal particles, soil pH and the cell-surface charge, which are discussed by Unc and Goss (2004). Differences in the hydrophobicity between vegetative and spore forms of some clostridial species have been observed (Wiencek, Klapes and Foegeding, 1990), possibly resulting in differences between how the cell forms adhere to soil particles and are retained in the soil.

The movement of water through a soil profile is strongly influenced by soil porosity, namely the pore micro- and macro-structure. Micropores of < 1-1.5 μm diameter may severely restrict the translocation of large, rod-

shaped bacteria in the soil (Mawdsley *et al*., 1995). Studies have indicated that an increased vertical translocation of bacteria can be expected in more macroporous soils, such as in structured clay, undisturbed, weathered or sandy soils (Mosaddeghi *et al*., 2009; Safadoust *et al*., 2011; Natsch *et al*., 1996). Soils with a predominately microporous structure, such as silty or sandy clays, are more efficient at filtering bacteria by physical obstruction, reducing vertical translocation of bacteria in percolating water (Safadoust *et al*., 2011; Mosaddeghi *et al*., 2009). Similarly, in mechanically disturbed soils (i.e. ploughed, tilled or repacked) the macrostructure is lost and vertical translocation through macropores is reduced (Safadoust *et al*., 2011; van Elsas, Trevors and van Overbeek, 1991). In saturated soils, preferential flow occurs through macropores and channels, hence increased vertical translocation of bacteria is observed due to the reduced filtering effect of the soil (Safadoust *et al*., 2011; Mawdsley *et al*., 1995; van Elsas, Trevors and van Overbeek, 1991). Gessler and Böhnel (2006) found evidence for vertical translocation of *C. botulinum* from upper to lower soil horizons after introduction of spore-contaminated compost on a loess soil. Newton *et al.* (2010) found a higher rate of recurrence of EGS associated with premises on sandy or loam soils, which could support the bacterial filtration mechanism detailed above. The importance of vertical translocation is emphasised after organic waste application. When pathogens are introduced to the top soil horizons, vertical translocation may be a key mechanism for dispersion of these pathogens to lower horizons, meaning contamination of crops or ingestion by livestock is less likely to occur.

2.5.2 Soil chemistry

There is very little research linking Clostridia behaviour in soils to any underlying soil chemistry. Very few significant associations have been identified, and some appear counter-intuitive and contradictory to other findings. For example, an association was identified between higher incidence of EGS with a higher soil nitrogen (N) content (Mccarthy *et al*., 2010). Clostridia are often diazotrophic organisms under anaerobic

conditions, and nitrogenase activity would be inhibited by increasing soil N, such as by ammonium or nitrate additions (Dorr de Quadros *et al*., 2012). Therefore, it is likely that the higher onset of EGS is due to increased pasture growth or pasture nitrogen content, and the resulting dietary change is the prevailing factor for illness. Dorr de Quadros *et al*. (2012) found a comparatively higher abundance of Clostridia in an oat/maize rotation without added N. As fertiliser addition is a common practice on both arable land and improved pasture, it is imperative that the resultant effect on clostridial populations is elucidated, particularly regarding EGS and other grazing-related clostridial diseases.

Many studies have found significant relationships between soil organic matter content and *C. botulinum* occurrence. *C. botulinum* Type A is more prevalent in soils with low organic matter content, whereas Types B and C show a strong association with higher organic contents (Dodds, 1992; Smith, 1978; Espelund and Klaveness, 2014). Böhnel and Lube (2008) postulate that the reduced amount of microporous aeration or raised nutrient content in decaying organic matter may assist in triggering *C. botulinum* growth.

2.5.3 Soil pH

Smith (1975b) identified a statistically significant relationship between higher counts of *C. botulinum* Types A and B in neutral to alkaline soils. However, in a later study, Smith (1978) found a higher prevalence of *C. botulinum* Type A in neutral-alkaline soils (average pH 7.5), whereas Type B prevailed in slightly more acidic soils (average pH 6.25). Environmental botulism outbreaks have also been associated with water that has a pH between 7.5–9 (Espelund and Klaveness, 2014).

C. perfringens has been isolated from soil with various levels of pH. Both *C. perfringens* and *Pn. sordellii* were prevalent across a wide range of soil pH in Costa Rican soils (del Mar Gamboa, Rodríguez and Vargas, 2005). This is coincident with other studies which have isolated both vegetative and spore forms of *C. perfringens* across a range of acidities (Smith, 1975;

Li, Sayeed and McClane, 2007; Stefanis *et al.*, 2014). *C. perfringens* was prevalent in the acidic soils (pH 4.5–6.5) surveyed by Voidarou *et al.* (2011). These studies are all in agreement with the laboratory-determined pH growth conditions (Table 2.1), and collectively indicate that soil pH is not a suitable predictor for *C. perfringens* and *Pn. sordellii* prevalence.

2.6 AGRICULTURAL MANAGEMENT

Studies have shown that different land management and land use practices can alter the bacterial biomass, diversity and community structure of the agricultural soils (Acosta-Martínez *et al*., 2008; Roesch *et al*., 2007). Changes to agricultural management could prove the most effective and realistic approach for reducing pathogen prevalence and abundance in soils, once the remedial or risk-reducing practices have been ascertained.

2.6.1 Tillage

Land owners are increasingly adopting sustainable cultivation management practices which minimise negative impacts of agriculture on the environment. One such technique, no-tillage, has been shown to increase the microbial diversity and biomass of soils (Ibekwe *et al*., 2002; Dorr de Quadros *et al*., 2012). Ibekwe *et al.* (2002) found that the total bacterial biomass and the diversity of ammonia-oxidizing bacteria were higher in no-till soils compared to conventional-till soils. More recently, Dorr de Quadros *et al.* (2012) explored soil microbial community change under no/conventional-tillage management using 16S rDNA analysis. The relative abundance of Firmicutes, represented by *C. beijerinckii*, showed a positive association with no-tillage systems. *Clostridium* species and other anaerobic bacteria were found to dominate no-tillage systems. It is likely that the reduced aeration in no-tillage soils favours the growth of anaerobic bacteria. Furthermore, the better soil structure may increase water-retention and enhance the formation of anoxic soil zones. Deep and

conventional-ploughing will bring unexposed and protected clostridial cells to the surface, where they may be inactivated by the sunlight (see section 2.4.3). Further research is needed to elucidate whether no-tillage systems favour the growth of pathogenic Clostridia. Increasing utilisation of this environmentally friendly technique could inadvertently raise the risk of clostridial disease.

2.6.2 Organic Soil Amendments

Historically, application of organic soil amendments (OSA) to agricultural land has been used as an efficient method of replenishing soil nutrients, improving soil quality and managing organic wastes (Venglovsky, Martinez and Placha, 2006). This traditional farming practice has become an increasingly attractive option due to heightened environmental and economic concerns regarding conventional chemical fertiliser use. Moreover, other bio-waste such as compost and anaerobic digestates are now frequently applied to soils. OSA such as manure, slurry, sewage and other bio-waste can contain a high pathogenic load, including many clostridial pathogens, which may be present in excreta or that survive/proliferate during the composting or digestion process (Bagge, Sahlström and Albihn, 2005; Sahlström *et al*., 2008; Bagge, Persson and Johansson, 2010; Torniainen *et al*., 2011). Clostridial pathogens have been observed and characterised in many different types of OSA (Table 2.2). Incorporating unsanitised organic wastes into agricultural soils can induce artificially high pathogen populations, increasing the risk of pathogen transmission into the food chain or to livestock (Girardin *et al*., 2005; Bagge, Persson and Johansson, 2010; Brochier *et al*., 2012). Furthermore, Clostridia have been shown to survive heat treatment processes, such as pasteurisation and thermophilic digestion, which are designed to remove or reduce pathogen loads (Bagge, Persson and Johansson, 2010; Olsen and Larsen, 1987; Sahlström *et al*., 2008). The application of OSA to agricultural land is governed by both European and individual member

Table 2.2 Examples of pathogenic Clostridia isolated from organic wastes

OSA	Species isolated	Method	Reference
Anaerobic digestates	*C. perfringens*	Culture based	Bagge, Sahlström and Albihn (2005)
Bio-compost	*C. botulinum*	Culture based + mouse bioassay	Böhnel and Lube (2008)
Bovine manure	*C. butyricum* + *C. perfringens*	Sequencing and R.E.[1] analysis of 16S rDNA	Ouwerkerk and Klieve (2001)
Dairy manure	*C. baratii, C. botulinum, C. butyricum, C. novyi, C. perfringens, Pn. sordellii* + *Pr. Bifermentans, C. sordelii*	Biochemical analysis + sequencing of 16S rDNA	Bagge, Persson and Johansson (2010)
Farmyard manure	*C. perfringens*	Culture based	Brochier *et al.* (2012)
MSW[2] compost	*C. perfringens*	Culture based	Brochier *et al.* (2012)
Sewage sludge	*C. perfringens*	Culture based	Dudley *et al.* (1980)
Swine manure	*C. butyricum*	DGGE of 16S rDNA	Leung and Topp (2001)

[1] Restriction enzyme analysis

[2] Municipal soil waste

state directives, which are detailed by Moynihan (2012). These legislative measures are designed to reduce the likelihood of pathogen transmission to crops and grazing livestock, and infiltration into groundwater. However, current legislation is aimed at minimising the survival of key pathogenic, non-spore forming bacteria such as *Escherichia coli* O157, *Salmonella* species and *Campylobacter* species. Endospore formation may further enhance the longevity of clostridial pathogens, meaning populations remain higher in the soil for a longer duration after introduction *via* OSA. Therefore, current legislation may provide insufficient controls on the transfer of clostridial pathogens from organic wastes to agricultural soils.

The behaviour and abundance of *C. perfringens* post-OSA application has been examined. Swine, fresh and aged cattle manure amendments induced significantly higher concentrations of *C. perfringens* in field run-off water after three simulated rainfall events (Thurston-Enriquez, Gilley and Eghball, 2005). Brochier *et al*. (2012) found no statistical difference between *C. perfringens* in amended and un-amended soils over a 33-month field study, despite two amendments (at 0 and 24 months) with concentrations of up to 1.5×10^4 CFU g^{-1}. They suggest that environmental stressors, particularly temperature, play a more pivotal role in the behaviour of *C. perfringens* in soil than does OSA application. Other pathogenic Clostridia which constitute part of human and livestock gut-flora, including *Cl. difficile* and *Pn. sordellii*, are more likely to be enhanced by OSA additions (Simango, 2006). By inference, species which are not common enteric microflora, such as *C. botulinum,* should not theoretically be increased by manure amendments. While some historical studies have found no positive correlation between manure or excreta application and increased *C. botulinum,* Dodds (1992) concludes that the localised, high incidence of some types of *C. botulinum* spore Types B, C and D reported in British soils, suggests that in some areas, animals can contribute to the incidence of *C. botulinum* in soils.

2.6.3 Grazing regime

It is important to differentiate intentional OSA to agricultural land from inputs of excreta from grazing animals. The former is usually a very controlled, legislated management process, whereby inputs can be sterilised and applied in known concentrations. The pathogen concentration in animal faecal matter is known to vary considerably based on the social, nutritional and immunological (i.e. age, stress, diet, disease) status of the animal (Waggett *et al*., 2010). Waggett *et al.* (2010) used culture and enzyme-linked immunosorbent assays (ELISA) to demonstrate the higher prevalence of *C. perfringens* in the faeces of EGS horses, than in healthy horses. *C. perfringens* and *Cl. difficile* are likely to be more prevalent in faeces of young foals than mature horses (Båverud, 2002; Tillotson *et al*., 2002; Newton *et al*., 2010), as well as many other young livestock (Songer, 1996). Kim *et al*. (2004) attributed the high prevalence of *C. perfringens* in various Korean soils to contamination by domestic animals. A case-control study indicated that the EGS incidence rate was significantly higher on recurrent grazing land, suggesting that soil or grass harbours pathogenic *C. botulinum* Type C. Interestingly, the same study found that co-grazing with ruminants reduced the risk of EGS in horses (Newton *et al*., 2010). Similarly, the incidence of blackleg in cattle (caused by *C. chauvoei*) is thought to increase year on year with heavily contaminated pastures (Bagge, Lewerin and Johansson, 2009; Hang'ombe *et al*., 2000). Simango (2006) isolated *Cl. difficile* in the faeces of various domesticated animals, with chicken faeces showing the highest prevalence (17%, n =115), which were thought to be the major source of soil contamination. It can be concluded that after clostridial pathogens are introduced to the soil *via* faecal matter, they can remain present in the soil in either spore or vegetative cell form. Additionally, mechanised manure removal, such as harrowing, could spread clostridial pathogens across a wider area of soil.

2.6.4 Cropping regime

It is known that crop type can affect soil microbial communities (Garbeva, van Veen and van Elsas, 2004). Carbon-rich root exudates and expansive root growth will alter the physicochemical environment of the rhizosphere and surrounding soil, thus affecting the soil microbial ecology. Bacterial-plant interactions are likely to be highly specific depending on the species or cultivar of crop used. Voidarou *et al*. (2011) found significant differences in the occurrence of *C. perfringens* spores and vegetative cells under 10 different cropping regimes, although this may be due to antimicrobial effects of pesticides applied to the crops or other parameters. Similar variations in *C. perfringens* prevalence under different cultivations have been documented in other studies (Stefanis *et al*., 2014). *C. perfringens* spores and vegetative cells occurred in 67% and 17% of samples respectively under maize cropping, yet both forms were absent under cabbage cropping. Again, it is unclear whether these variations can be attributed to the cropping regime alone, which emphasises the need for more controlled experiments where confounding variables can be minimised.

2.7 METHODS OF DETECTION

A more comprehensive understanding of clostridial prevalence, abundance and behaviour in soils can be obtained with extensive, geographically-widespread soil-surveys covering the complete range of soil types, agricultural regimes and climates. This, in combination with laboratory microcosm experiments that manipulate the key environmental and physicochemical factors mentioned, would help elucidate the complex behaviour of Clostridia in soil. However, what underpins the validity, accuracy and applicableness of data from future studies is the use of suitable diagnostic techniques which facilitate; 1) the generation of accurate pathogen prevalence and abundance data, 2) the differentiation between vegetative growth and endospore cell forms and 3) easy

reproducibility and application in microbiology laboratories. This section explores the common techniques used in clostridial diagnostic work and critiques their suitability for future studies.

2.7.1 Culture-based assays

Prior to the development of molecular based techniques, clostridial enumeration and identification was performed using culture, biochemical and immunological-based assays. Culture assays can be used to identify and enumerate some clostridial species in samples using the plate count or the most probable number methods. Samples are also frequently cultured in enrichment media prior to other diagnostic tests. Environmental samples are typically cultured on growth medium such as cooked-meat medium (Sonnabend, Sonnabend and Krech, 1987), trypticase-based medium (Burns, 2011; Brüggemann *et al*., 2004; Meng *et al*., 1999), reinforced clostridial medium (Stringer *et al*., 2013; Khanna, Goyal and Moholkar, 2013), or brain heart infusion medium (Smith, 1975; Nicholson, Groves and Chambers, 2005; Li, Sayeed and McClane, 2007) under anaerobic conditions. Media are regularly reformulated to provide better growth conditions, or to allow improved differentiation of species. Culture-based identification and enumeration is still a widely used technique (Vijayavel and Kashian, 2014; Sonnabend, Sonnabend and Krech, 1987). One benefit is that high temperature short time pasteurisation (typically 70°C for 2-10 minutes prior to incubation) can be applied to destroy all vegetative cells, leaving just spore-forms. This provides a simple but effective method for discriminating between cell-growth forms. Although cost effective, culture-based assays face several limitations. Only a portion of clostridial cells can be cultured, those that are injured will not be detected, although may still be capable of pathogenesis. Furthermore, atypical colonial morphology can lead to misidentification in culture-based assays. Selective or differential growth media have not been developed for many clostridial species, meaning discrimination between some closely related species in co-contaminated samples is impossible.

2.7.2 Biochemical assays

Traditional biochemical identification of anaerobic bacteria comprises a complicated, multi-step methodology typically culminating in the analysis of metabolic end-products using gas-liquid chromatography (Burlage and Ellner, 1985; Perry, 1985). Biochemical assays allow the definitive identification of anaerobic bacteria, opposed to just the presumptive identification using culture-based approaches (Head and Ratnam, 1988). However, the slow growth of obligate anaerobic bacteria in comparison to aerobic or facultative bacteria delays identification. A range of biochemical kits, such as PRAS II (Burlage and Ellner, 1985), have been developed to identify anaerobic bacteria such as Clostridia. These kits consist of a range of pre-reduced biochemical substrates in which isolated colonies are identified and semi-quantified. A range of non-growth dependent biochemical kits are also available, utilising constitutive enzymes on specific substrates; many of which have chromogenic reaction products (Burlage and Ellner, 1985). Kits used by previously cited studies include RapID-ANA (Hang'ombe *et al*., 2000) and API systems (Cordoba *et al*., 2001; Kim *et al*., 2004; Lindström *et al*., 2001). Despite their rapidity, studies have shown that these biochemical kits lack both the sensitivity and specificity to accurately identify some pathogenic Clostridia*,* such as *C. botulinum* (Lindström *et al*., 1999) and *Cl. difficile* (Head and Ratnam, 1988)*.* Biochemical analyses are typically laborious, expensive and time-consuming (Keto-Timonen *et al*., 2006).

2.7.3 Immunological assays

Immunological assays are readily used for identification of *Clostridium* pathogens, or commonly for the detection of specific toxins. The gold standard method of detection of clostridial toxins, such as BoNTs, was typically a microbiological method combined with the *in vivo* standard mouse bioassay (SMB) (Fenicia *et al*., 2011). The SMB procedure is well-documented in previously cited papers (Meng *et al*., 1999; Fach *et al*., 2009; Böhnel and Lube, 2008). Although sensitive and highly specific, the

approach is expensive, time-consuming and now discouraged in a number of jurisdictions for ethical reasons regarding animal experimentation (De Medici *et al*., 2009; Fenicia *et al*., 2011; Cordoba *et al*., 2001; Lindström *et al*., 2001).

Enzyme immunoassay (EIA) diagnostic methods are a popular approach for toxin detection, such as ELISA, which is used in many of the studies previously cited. This assay, and variations of, are used to detect and quantitate specific antigens, such as clostridial toxins, with a good overview of the methods given by Paulie *et al.* (2001). EIAs have proven to be sensitive, high-throughput assays for toxin detection (Paulie, Perlmann and Perlmann, 2006; Peterson *et al*., 2011). Another related approach is using immuno-fluorescent assays (Hang'ombe *et al*., 2000; Kim *et al*., 2004), in which fluorophore-labelled antibodies attach to antigens of interest, such as clostridial toxins. These methods have major drawbacks in the identification and enumeration of bacterial cells. Different target pathogens can produce identical proteinaceous toxins (such as the α, β, γ and δ toxins of *C. haemolyticum, C. chauvoei* and *C. septicum* (Hatheway, 1990)), which are indistinguishable using ELISA. Furthermore, the technique can only be used indirectly to infer bacteria populations, as some bacteria will not produce toxins, and the true pathogen abundance may be underestimated. As some pathogens have multiple toxinotypes, separate tests would need to be conducted for each possible toxinotype, increasing the cost and complexity of the experiment. Other immunological diagnostic kits utilise the same scientific principles as ELISA, such as reversed passive latex agglutination test kits, and have been shown to be useful for toxin identification, although the same limitations apply, and poor sensitivity and specificity in comparison to the SMB have been documented (Peterson *et al*., 2011; Head and Ratnam, 1988).

2.8 MOLECULAR TECHNIQUES

In recent decades, huge advances have been made in the development of molecular diagnostic tools. These techniques, in addition to constantly-growing genetic databases, allow increasingly fast and sensitive bacterial strain detection and characterisation (Cordoba *et al*., 2001). This section will critique some of the important molecular methods that have or could be used for detection, enumeration and study of clostridial pathogens in the soil.

2.8.1 PCR-based techniques

Most of the research discussed in this review utilises polymerase chain-reaction (PCR) based techniques. Not only does PCR permit microbial analysis on relatively small or rare environmental samples, but it has provided the basis for a more thorough and accurate taxonomic assessment of the Clostridia class (Collins *et al*., 1994). The first step requires extraction of bacterial DNA from the environmental sample. A range of high-efficiency kits are available to facilitate this task. Depending on the aim of the study, specific oligonucleotide primer pairs are designed to target specific nucleotide regions on the target-organism(s). Primers are designed with *in silico* analysis of predetermined nucleotide sequence libraries using appropriate software. The protocol can be designed to target highly-specific genetic loci, such as toxin-coding genes, or highly-conserved regions such as the 16S rDNA gene. After PCR, successful amplification of the gene fragment would indicate the presence of the particular pathogen strain in the sample (Lindström *et al*., 2001; Fach *et al*., 2009). Techniques such as multiplex, nested and semi-nested PCR employ multiple primers or primer pairs, enabling simultaneous and specific detection of different target genes, species or toxinotypes (Lindström *et al*., 2001; De Medici *et al*., 2009). A further development of this approach is to use fluorescently-labelled primers (probes) that allow real-time visualisation of fragment amplification during PCR, through real-time quantitative PCR (RT-qPCR), or the automated sizing of amplicons,

such as with automated ribosomal intergenic spacer analysis (ARISA) or terminal-restriction fragment length polymorphism (T-RFLP). These approaches allow a sensitive and quantitative approach to microbial diagnostics, and therefore can be used as an alternative to the SMB or other toxin-detecting immunological assays (Fach *et al*., 2009; Fenicia *et al*., 2011). Many authors have demonstrated the usefulness and sensitivity of using PCR protocols to detect clostridial pathogens in clinical (Donskey *et al*., 2014; Gurtler, Wilson and Mayall, 1991; Song, Liu and Finegold, 2004), food (Cordoba *et al*., 2001; Lindström *et al*., 2001) and environmental samples (Fenicia *et al*., 2011; De Medici *et al*., 2009; Fach *et al*., 2009; Hielm *et al*., 1996).

Broad-range PCR primers can be used to target multiple species or toxinotypes. Various downstream molecular fingerprinting techniques can then be used to differentiate between the species/strains using one protocol. In this instance, primers are normally designed to target the 16S or 23S rRNA gene regions, due to their ubiquity among bacteria and large, pre-existing gene libraries available to analyse. Methods such as denaturing gradient gel electrophoresis (DGGE), temperature gradient gel electrophoresis (TGGE) (Marzorati *et al*., 2008), amplified fragment length polymorphism (AFLP) (Keto-Timonen *et al*., 2006), ARISA (Dahllöf, 2002; Feligini *et al*., 2015), T-RFLP (Khoruts *et al*., 2010), single strand conformation polymorphism (SSCP) (Marzorati *et al*., 2008; Smalla *et al*., 2007) and other similar techniques have been utilised as diagnostic tools for Clostridia detection. One common question regarding all the aforementioned techniques is whether they are specific enough to allow the differentiation between closely-related targets with a high genetic similarity. Keto-Timonen *et al.* (2006) used AFLP to successfully differentiate between 22 out of 24 different clostridial species including some toxic strains. T-RFLP has not currently been used to identify different clostridial strains or species from individual samples, although differentiation to cluster level has been achieved (Miyamoto, Kawahara and Minamisawa, 2004).

PCR-based approaches have a high applicableness for pathogen detection, although the methods are not without their limitations. Wintzingerode *et al.* (1997) detail the common pitfalls associated with soil microbial analysis using PCR-based approaches. Erroneous results can arise from inefficiencies during cell lysis and nucleic acid extraction stages. Certain soil components, such as humic and fulvic acids are known to inhibit PCR amplification, although this can be rectified by use of additives such as non-acetylated bovine serum albumin (BSA). Such biases may lead to results which are not truly quantitative (Acosta-Martínez *et al*., 2008). Another notable problem with regards to studying soil microbiology using PCR-based techniques is due to the exceptionally high and poorly understood soil microbial diversity (Tiedje *et al*., 1999). In what is commonly referred to as the 'black-box' of soil ecology, it is unknown as to what effect the many undiscovered species could have on interpreting PCR-based data. Many rare species may remain undetected, and unknown species showing high genetic homology could lead to the generation of false-positives (Culman *et al*., 2008).

2.8.2 Sequencing

Nucleotide sequencing is now the standard technique for the confirmative detection of pathogens. This tool also provides an increasingly cost-effective way to identify, survey and compare bacterial communities across different environments (Burke and Darling, 2014). In short-read sequencing, typically a conserved region of DNA is amplified using PCR, and the amplicons excised. Gene fragments are then sequenced using a number of techniques, such as pyrosequencing (Acosta-Martínez *et al*., 2008; Roesch *et al*., 2007) or Illumina sequencing (Burke and Darling, 2014; Dorr de Quadros *et al*., 2012). The sequences can then be compared to known species sequences on existing databases for identification and genetic comparison with other strains and are deposited on databases for future use. While identification of multiple species from one sample is technically feasible, the method then becomes more expensive and

requires considerable *in silico* analysis to identify sequence reads. Whole genome sequencing (WGS) is now considered the ultimate tool for isolate identification and genetic analysis (Salipante *et al*., 2015). The constantly falling cost of sequencing (Burke and Darling, 2014) and the comprehensive range of *in silico* analytical software make this approach increasingly suitable with regards to identifying clostridial pathogens in the soil. However, the cost may still be prohibitive for large pathogen-surveillance studies, and suitable media are still needed to isolate and differentiate different species. WGS sequencing would be the most sensitive, informative approach to generate important data on genetic and function diversity of Clostridia in the soil. However, a high-throughput, inexpensive molecular assay to screen for pathogens (such as T-RFLP, ARISA), would allow for a comprehensive survey of agricultural soils. Contaminated soils can be identified, pathogenic strains isolated, and WGS used to elucidate how genotypic characteristics allow adaption to the soil environment and changes in virulence. An overview of a suitable methodology is shown in Figure 2.1.

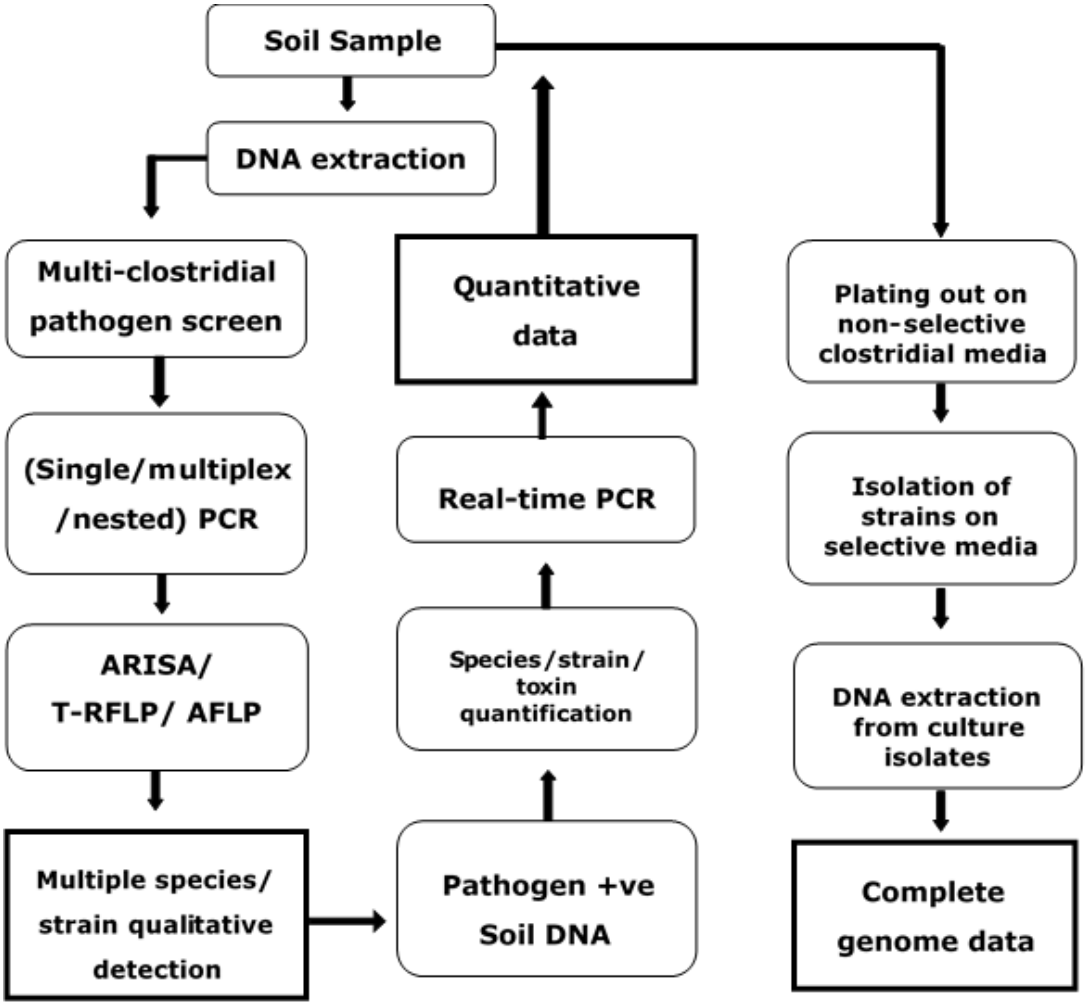


Figure 2.1 Schematic demonstrating a suitable workflow for the detection, quantification and confirmative identification of clostridial species, strains and toxins in the soil. Boxes indicate the potential outputs of specific procedures.

2.9 CONCLUSIONS

Clostridial pathogens are aetiological agents for many significant human and veterinary diseases. The soil is the main reservoir for most key pathogens, such as *C. botulinum, C. tetani* and *C. perfringens.* Studies have demonstrated the prevalence of these pathogens, as either vegetative cells or as spores, in most soil environments including

agricultural soils. Quantifying the prevalence and abundance of clostridial pathogens in agricultural soils is of key importance, as this environment represents the first critical control point in the food contamination pathway, and route of infection for susceptible grazing animals (Haagsma, 1991; Stefanis *et al*., 2014). There have been numerous studies attempting to identify the occurrence of key pathogenic species and toxinotypes in farm soils, although the findings of these studies differ significantly. This may reflect the high variability in prevalence and abundance, or the differences in analytical methods used. However, what many studies acknowledge is the need for future research to understand the factors which affect Clostridia behaviour in soil (Stefanis *et al*., 2014; Newton *et al*., 2010; Espelund and Klaveness, 2014). Existing soil pathogenic indicator species are not suitable for predicting Clostridia behaviour due to the environmental persistence of spores in the soil, and the highly variable physiologies shown between species and strains. Some of the environmental, soil physicochemical and agricultural management factors which are likely to influence pathogen survival were discussed. However, only a tentative link can be made between these factors and their effect on clostridial growth in soil. Future research should aim to elucidate the role of these and other factors on pathogen behaviour, using both microcosm and field studies. This will allow the improved pathogen modelling and mapping, development of better risk management strategies, with an aim to reducing the incidence of clostridial disease.

Fundamental to the validity of future research is the adoption of appropriate, sensitive and selective techniques that allow the differentiation between vegetative and spore growth form and clinically different species and strains. Suitable methodologies will allow a high-throughput, cost-effective and widely accessible application. A variety of suitable approaches were identified, although the falling cost and sensitivity of PCR and sequencing techniques makes them ideal tools for clostridial soil diagnostics. In particular, ARISA, T-RFLP and RT-qPCR, in conjunction with culture and sequencing-based approaches, are the most

appropriate technologies for highly sensitive, high-throughput methods for multiple-species and strain identification, enumeration and even toxin detection with the use of suitably designed primers.

CHAPTER 3: A REVIEW OF CLOSTRIDIAL PHYLOGENY USING MULTILOCUS SEQUENCE ANALYSIS

3.1 ABSTRACT

The genus *Clostridium* is a phenotypically and genetically diverse taxon containing many pathogenic and biotechnologically important species. Taxonomic assignment based on phenotypic characteristics or 16S rDNA alone is not sufficient to infer the evolutionary relationships between strains and species. In this study, clostridial phylogenies were inferred using multilocus sequence analysis. Nucleotide coding sequences for conserved housekeeping genes were used to produce maximum-likelihood and Bayesian phylogenies for 101 clostridial strains. These phylogenies support the recent removal of species to alternative families within the class Clostridia, including the reassignment of pathogens *C. bifermentans, C. difficile* and *C. sordellii* to the family *Peptostreptococcaceae*. The phylogenies allowed discrimination between closely-related species and strains, identifying the likely misclassification of some organisms. *Clostridium aceticum* fell outside the well-defined clade of *Clostridium sensu stricto,* and numerous strains appear to have been previously misnamed. Tree comparison metrics were used to identify individual gene trees which had the most similar topology to that of the multigene tree. The *inf*B gene had the most similar topology, and phylogenies were inferred from this gene for 238 *Clostridium* strains. Species groupings were consistent with the multigene tree, while numerous examples of taxonomic conflicts, including for important pathogenic species were identified. Rapid and accurate identification of clostridial isolates would be possible by sequencing the *inf*B gene and using the derived sequences for phylogenetic reconstruction. This method could be a suitable precursor to whole genome sequencing.

Keywords; *Clostridium*, Clostridia, taxonomy, phylogeny, bioinformatics.

3.2 INTRODUCTION

The genus *Clostridium* contains a diverse assemblage of bacterial species, many of which have significant medical, environmental and biotechnological importance. *Clostridium* species were originally designated to this genus based on phenotypic and biochemical traits, such as being Gram-positive, rod-shaped anaerobic bacteria capable of endospore formation (Hatheway, 1990; Andreesen, Bahl and Gottschalk, 1989). Recent phylogenetic analysis using DNA sequences has cast serious doubt on the classification of species based on phenotypic characteristics alone, while early DNA-based phylogenies may also be inaccurate. Nomenclative and taxonomic discontinuities exist in many clostridial species across key resources. This has ramifications for analysis, and databases, and in practical terms this could result in incomplete sampling efforts or delays in the identification of novel bacterial species and strains (Pritchard *et al*., 2016). Accurate identification and classification is particularly important for clinical isolates, so that practitioners can be rapidly directed to the relevant literature on pathogen epidemics, management and treatment options, such as disease reporting or quarantine actions. This therefore justifies the need for a new, standardised phylogenetic approach that allows a more robust approximation of the likely evolutionary history, and more precise taxonomic assignment for existing and new clostridial species.

The *Clostridium* genus first underwent significant taxonomic revision due to advances in sequencing technologies. Studies by Collins *et al.* (1994) and Stackebrandt *et al.,* (1999) utilised 16S ribosomal RNA (rRNA) gene typing to reassign many *Clostridium* species to novel or alternative genera within the family *Clostridiaceae,* or alternative families in the larger class Clostridia (Collins *et al*., 1994; Stackebrandt *et al*., 1999; Weigel, Tanner and Rainey, 2006). The current consensus is that only species within 16S rDNA clade Cluster I should be considered true representatives of the *Clostridium* genus (*Clostridium sensu stricto*), for which *C. butyricum* is the

current type species (Gupta and Gao, 2009; Collins *et al*., 1994; Stackebrandt *et al*., 1999). This cluster was found to be a genetically heterogeneous group demonstrating markedly different phenotypes. This genus represents many important pathogenic species, although some pathogens, including *C. difficile* and *C. sordellii*, were reassigned to the family *Peptostreptococcaceae* (Ludwig, Schleifer and Whitman, 2009; Yutin and Galperin, 2013), although many resources still have not adopted the updated taxonomy. Although 16S rRNA typing is a useful and well-established method for phylogenetic studies, the highly conserved nature of this gene means that clinically different pathotypes, and even certain species are indistinguishable from one another. For example, the 16S rDNA sequence difference between the pathogens *C. chauvoei* and *C. septicum* is less than the commonly used operational definition of a species for the 16S rRNA gene (Kuhnert *et al*., 1996; Nguyen *et al*., 2016). Multiple copies of the 16S rRNA gene within a genome can exhibit divergent 16S rDNA sequences (Cilia, Lafay and Christen, 1996; Dhalluin *et al*., 2003; Christensen *et al*., 2004). The sequence divergence can exceed this operational threshold, resulting in sequences from the same organism being misclassified as more than one species (Pei *et al*., 2010). There are also issues associated with reconstructing phylogenies from individual genes (Daubin, Gouy and Perrière, 2002; Gadagkar, Rosenberg and Kumar, 2005). Such problems include the occurrence of paralogous genes from duplication events, gene recombination events and horizontal transfer between species or strains, all of which can lead to incorrect tree inference (Lang, Darling and Eisen, 2013; Janda and Abbott, 2007). Furthermore, single-gene analyses rely upon the evolutionary history of the chosen gene reflecting that of the whole organism, which is not always the case. Some genes can be transferred to other bacterial species *via* mobile genetic elements, such as plasmids, making them unsuitable for evolutionary inference for the host organism. While plasmids often contain toxin and antibiotic resistance-encoding genes, plasmids have been found to seldom carry essential housekeeping (HK) genes (Tazzyman and Bonhoeffer,

2014), making these HK genes preferable candidates for phylogenetic analysis.

Many HK genes are also universal to all bacteria, with high evolutionary conservation and low selection pressure making them suitable markers for evolutionary comparison and inference between distantly-related bacteria (Christensen *et al*., 2004). Horizontal gene transfer seldom occurs in HK genes (Sawada *et al*., 1999), making them a good index for determining genome evolution between species. Such genes have been used for phylogenetic studies for *Aeromonas* (Soler *et al*., 2004), *Pasteurellaceae* (Christensen *et al*., 2004), *Pseudomonas* (Mulet, Lalucat and García-Valdés, 2010) and *Bacillius* (Liu *et al*., 2013) genera. Recently, attempts to better define *Clostridium* phylogeny have utilised HK genes other than 16S rDNA (Gupta and Gao, 2009; Yutin and Galperin, 2013). Studies by Gupta and Gao (2009) and Yutin and Galperin (2013) used 37 and 50 highly conserved proteins respectively, including mainly ribosomal proteins, to better understand relationships between selected clostridial species. Yutin and Galperin (2013) proposed reassigning 78 previously-assigned *Clostridium* species to six new genera, although the proposed changes in taxonomy were not adopted as they did not comply with the Bacteriological Code and were not validly published in the International Journal of Systematic and Evolutionary Microbiology (Lawson *et al*., 2016). However, they suggest that informational protein-coding sequences will provide a valid tool for phylogenetic inference and can successfully complement rDNA-based trees in building the ultimate genome-based classification system for Bacteria and Archaea. Utilising multilocus phylogenies can increase the discrimination power between species and strains (Soler *et al*., 2004; Liu *et al*., 2013; Mulet, Lalucat and García-Valdés, 2010), and reduce systematic biases that could be associated with individual gene inference (Tambong *et al*., 2014).

In this study, conserved, non-ribosomal HK gene nucleotide coding sequences (NCS) were comprehensively analysed and used to generate a

high-resolution phylogeny of 101 *Clostridium* and former-*Clostridium* species. This included many recently-sequenced complete and draft genomic sequences which have not yet been used for phylogenetic inference. Multiple strains were used for some species, where available, to enhance knowledge of intraspecies genetic relationships, including that of different toxinotypes and clinical strains from important clostridial pathogens. This approach facilitated greater discrimination between strains and species, circumventing the issues associated with using just the 16S rDNA. Comprehensive checks for gene paralogy and tree incongruence were performed throughout, and different methods of phylogenetic reconstruction were compared to further characterise the results. Various methods were used to identify reliable genetic markers for phylogenetic inference and strain differentiation, with the *inf*B gene proving to be the best candidate. This gene was then used to infer the phylogeny of 238 strains from a wider pool of *Clostridium* genomes.

3.3 METHODS

3.3.1 Taxa and gene selection

Accession numbers for all publicly available complete genome coding sequences for *Clostridium* spp. were retrieved from the NCBI Genome database [http://www.ncbi.nlm.nih.gov/genome] using the search terms *Clostridium* (or former *Clostridium* species denoted by [*Clostridium*]) (accession numbers retrieved on 09.09.17, see Supplementary Table 1). NCBI RefSeq sequences were preferentially chosen where available, as these sequences are annotated and non-redundant (O'Leary *et al*., 2016). Two hundred and six essential housekeeping genes ubiquitous to all bacteria were initially selected for analysis, based on the 'minimal bacterial gene set' identified by Gil *et al*. (2004). NCS of <750 base pairs in length were excluded, so that the length of NCS to be analysed could be maximised using a lower number of genes.

3.3.2 Sequences retrieval and preparation

Datasets were created for each gene by implementing the BLASTN (Altschul, 1997) tool, using a gene NCS from a randomly selected taxa from our database as a template to retrieve orthologous homologous sequences from the other selected genomes. This was achieved using the 'align two or more sequences' function and the 'discontiguous megablast' algorithm against the RefSeq Genome database. Aligned hits for each taxa with >95% identity and a $1e^{-7}$ expected value value cut-off were retrieved, following a similar method documented in other studies (Fitzpatrick *et al*., 2006). Some of the genes included had multiple gene copies in each genome. Although the use of single copy genes is generally applied in most phylogenetic studies (Soler *et al*., 2004; Lang, Darling and Eisen, 2013; Daubin, Gouy and Perrière, 2002), this issue was circumvented by selecting the gene that shared the highest sequence identity to the search template. This allowed for analysis of a wider set of genes. Gene sets containing paralogous sequences, or where sequences could not be retrieved for all taxa were excluded from the analysis.

Sequences were renamed, reverse complemented if necessary and visually inspected using Mesquite v.3.04 (Maddison and Maddison, 2001). The sorted sequences were then aligned using MUSCLE (Edgar, 2004) with 16 iterations, trimmed and filtered to contain positions with less than 50% gap characters using CLC sequence viewer software v7.6.1 (CLC Bio, Cambridge, MA). Gene NCS which could not be aligned were excluded from analysis. Gene alignments were then exported in FASTA format for further analysis, or converted to relaxed PHYLIP format for phylogenetic analysis using *R* v3.4.3 (Buchanan, 1993).

3.3.3 Analysis of sequence data

Summary statistics for gene sequence alignments were computed using DnaSP version 6.11.01 (Rozas *et al*., 2003). Alignment length, the G + C content, number of haplotypes (h), number of polymorphic sites (S) and

nucleotide diversity per site (π) were all calculated with Jukes–Cantor correction statistics. Tajima's D and Fu and Li's D* and F* neutrality tests were also performed to ensure sequences evolved neutrally (Fu and Li, 1993; Tajima, 1989). The alignments were concatenated to create one alignment file using SequenceMatrix v1.7.8 software (Vaidya, Lohman and Meier, 2011). The optimal partitioning scheme and the corresponding evolutionary rates for each partition were estimated using PartitionFinder v2 program (Lanfear *et al*., 2016) written in Python 2.7. To determine the best partitioning scheme, a starting tree was estimated from the data with using Maximum-Likelihood (ML) with RAxML (Stamatakis, 2014; Stamatakis, Hoover and Rougemont, 2008), with 'relaxed clustering algorithm' applied and best model selection based on Akaike Information Criteria (AICc) score.

ML distances matrices (DM) were created from sequence alignments for each gene, each partition and the concatenated alignment using the appropriate evolutionary model with PAUP v.4.0a software (Swofford, 1998). Prior to running phylogenetic models, congruence among distance matrices (CADM) was assessed using the CADM.global and CADM.post test functions within the APE package (Paradis, Claude and Strimmer, 2004) and implemented in *R*. This was used to indicate partitions that show a different evolutionary history and may skew the modelled phylogeny. Briefly, the CADM functions are used to generate Kendall's W concordance statistic, with the null hypothesis that two DM are completely incongruent (Legendre and Lapointe, 2004; Campbell, Legendre and Lapointe, 2009; Tambong *et al*., 2014). When there is some level of congruence between DM, *a posteriori* CADM tests are used to determine the contribution of each matrix to the overall congruence (Campbell, Legendre and Lapointe, 2009). For *a posteriori* tests, the Holm and the more conservative Bonferroni P-value corrections were applied, with 999 permutations, and incongruent ($p > 0.05$) partitions were removed from the concatenated alignment.

To determine the discriminative power for selected genes, pairwise least square tendency lines were generated between all genes and the concatenated alignment. To achieve this, DM were imported to *R* and unfolded using the melt function from the reshape2 package (Wickham, 2007). Duplicates were removed, before creating a data frame of pairwise distances between each gene and the concatenated sequence. Linear regressions were then implemented in *R* to derive the slope of the tendency line, and the correlation coefficients (R^2). The discriminatory power of each gene was calculated as the ratio between its slope and the concatenated 1/1 slope (Mulet, Lalucat and García-Valdés, 2010; Tambong *et al*., 2014).

3.3.4 Phylogenetic analysis

Phylogenetic analysis was performed on the concatenated alignment using ML and Bayesian Inference (BI) methods. The concatenated alignment was also uploaded to the CIPRES scientific gateway (Miller, Pfeiffer and Schwartz, 2010) so that up to 48 cores of processing power could be utilised, reducing the overall modelling time. ML inference was performed using the RAxML-HPC2 v8.2.10 (Stamatakis, 2014; Stamatakis, Hoover and Rougemont, 2008) software on CIPRES, with the *–f a* option used to generate 1000 rapid bootstrap replicates with the appropriate |3evolutionary model and 25 distinct rate categories. A file was also supplied designating the best partitioning scheme as determined previously. ML trees were also generated for individual genes, using the correct evolutionary model and the same parameters as for the concatenated tree. BI analysis was performed using MrBayes v.3.2.6 (Ronquist *et al*., 2012) on CIPRES. The appropriate evolutionary model was used. The Markov chain Monte Carlo was run with a temperature parameter of 0.2 for 1×10^6 generations, with sampling InL values and trees at 1000-generation intervals, swap frequency of 3 per generation, and 2 independent runs with 1 cold and three hot chains. The same partitioning scheme was also supplied. Bayesian posterior probabilities were calculated.

3.3.5 Comparison of trees

Unrooted phylogenetic trees were imported in Newick format to *R.* Three distance metrics were used to compare differences between concatenated ML and BI trees and individual partition ML trees. A review of different metrics indicated that the unweighted Robinson–Foulds (RF) (Robinson and Foulds, 1981) and the weighted-RF metrics performed well at comparing phylogenies (Kuhner and Yamato, 2014). The unweighted RF method compares trees based on topology alone, whereas the weighted RF considers differences in topology and branch length. A third method uses a Bayesian modelling approach to calculate the minimum number of subtree prune-and-regraft (SPR) operations required to resolve inconsistencies between two competing unrooted trees (de Oliveira Martins, Leal and Kishino, 2008), as well as the path difference metric of Steel and Penny (Steel and Penny, 1993). All methods were implemented by treedist function of the Phangorn package (Schliep, 2011), and visualised as heatmaps using the heatmap.2 package in *R.*

3.3.6 Gene-based phylogenies

Genes (or partitions) which showed the most similar phylogenies to the concatenated trees and demonstrated similar or better discriminative power were used to produce single-gene phylogenies of *Clostridium sensu stricto* species. Briefly, a random NCS from the original dataset was used as a query for a discontiguous megablast against the RefSeq Genome and whole-genome shotgun contigs databases and aligned hits (with highest identity and query coverage) were retrieved. When possible, at least three alignment sequences were retrieved for each species, particularly for pathogenic strains or strains known to have multiple toxinotypes. These sequences were prepared as described previously and used to create ML phylogenetic trees. All trees were visualised using Interactive Tree of Life v3 software (Letunic and Bork, 2016).

3.4 RESULTS

3.4.1 Taxa and genes used for analysis

A search of the NCBI Genome database retrieved accession numbers for 101 complete clostridial genome sequences, including 58 current *Clostridium sensu stricto* strains from 23 validly-described species, with the remaining strains belonging to 18 species formerly classified as *Clostridium*. For species and strain details see Supplementary Table 1. From the initial 206 housekeeping genes considered for analysis, 50 genes were selected that were >750 nucleotide bases in length and produced good quality visual-alignments. These genes are listed in Supplementary Table S2. Summary statistics were generated for each gene, shown in Supplementary Table S2. Gene alignments ranged in length from 746 – 2487 bp (after trimming). G + C contents were low for all genes, ranging from 28.2-37.6%, while the number of haplotypes remained constant between alignments (57-74). The percent of polymorphic sites was more varied, ranging from 50 - 80% of total sites, while nucleotide diversity per site ranged from 0.23 - 0.37 (Supplementary Table 2). Neutrality tests for all genes were non-significant ($p > 0.05$).

The alignment sequences for the 50 genes were concatenated to give a 69,826 bp sequence alignment. PartitionFinder v2 software was used to calculate the appropriate partitioning scheme and correct evolutionary models for phylogenetic analysis. An optimal partitioning scheme of 42 partitions was devised, grouping {*asp*S, *prf*A, *phe*S}, {*dna*K, *gro*EL}, {*eng*A, *ych*F}, {era, *ionA}*, {*gyr*A, *nus*A}, {*ksg*A, *mra*W} and {*obg*E, thrS} together. All other genes remained as separate partitions. The generalised time-reversible (GTR) plus gamma distribution (G) and proportion of invariable sites (I) was found to be the best evolutionary model for all partitions and was used in phylogenetic analysis and to generate all DM. DM for each partition were assessed for incongruence with all other partition DM using the CADM.post function in *R.* For every partition comparison, p= <0.05 (with both Holm and Bonferroni P value corrections

[data not shown]), meaning that the null hypothesis of complete incongruence was rejected, and all partitions were included in the final concatenation.

3.4.2 Phylogenetic analysis

Phylogenetic trees were generated for all individual genes using ML and for the concatenated sequence using both ML and BI methods on the CIPRES platform. ML and BI trees based on the concatenated sequence alignment for 50 genes are shown in Figures 3.1 and 3.2 respectively. ML and BI methods produced highly congruent tree topologies, both showing monophyletic groups for *Clostridium sensu stricto*, *Ruminiclostridium* and *Lachnoclostridium* with 100% bootstrap or Bayesian posterior probability support. All species contained within the large *Clostridium* clade (green area in Figures 3.1 and 3.2) are in accordance with the *Clostridium sensu stricto* species identified in Weigel *et al*., (2006) and Gupta and Gao (2009). Both phylogenies group the *Clostridium* main clade into two distinct subclades containing identical species. *C. botulinum* Group I (proteolytic) and other Group III strains are in a different subclade from *C. botulinum* Group II strains, with individual strain clustering highly consistent between ML and BI trees. The subclade containing *C. botulinum* Group I also contain acetogenic and solvent-producing Clostridia of commercial importance. Both models split the 22 *Clostridioides difficile strains* into four separate lineages, supported by 100% bootstrapping values (Figure 3.3 A).

The phylogenies also indicate the previous misnaming of some bacterial strains. *C. pasteurianum* str. NRRL B-598 groups with all *C. beijerinckii* stains (Figures 3.1 and 3.2), while the toxigenic *C. botulinum* str. Prevot 594 and *C. botulinum* str. Osaka05 group with the non-pathogenic *C. sporogenes* strains (Figure 3.3 B). Additionally, both phylogenies showed very little genetic distinction between *C. autoethanogenum* and *C. ljungdahlii*. Sequence analysis of the two 50-gene sequences showed that they shared a 98.8% sequence identity. *C. aceticum* is the only species

that is currently named as a *Clostridium* species (as of 09.2018) that did not fall within the *Clostridium sensu stricto* clade. The nearest relatives to *C. aceticum* were *Gottschalkia acidurici* and *Acetoanaerobium sticklandii*. A BLASTN of the 50-gene concatenated *C. aceticum* sequence against the nucleotide collection database revealed closest matches were *Alkaliphilus oremlandii* (CP000853.1) and *Alkaliphilus metalliredigens* (CP000724.1).

3.4.3 Comparison of trees

ML trees were generated for all 42 partitions, and tree similarity was compared between each partition, ML and BI trees using RF, wRF, SRP and path difference distance metrics. The results are shown in Figures 3.4 A-D, with blue colours indicating lower metric hence more similar topologies and/or branch-lengths, ranging to red which indicates higher levels of incongruence between trees. ML and BI tree comparisons gave the lowest scores for all metrics, although scores were omitted from heatmaps to improve clarity. Results between the four metrics were variable. A core of gene partition trees showed the smallest difference in topology from concatenated trees and each other. This is most evident in the RF, wRF and path distance heatmaps (Figure 3.4 A, B and D). This core includes {*aspS, prfA, pheS*}, *inf*B, *ffh, pfkA*, {*obgE, thrS*} and *secA* partition trees.

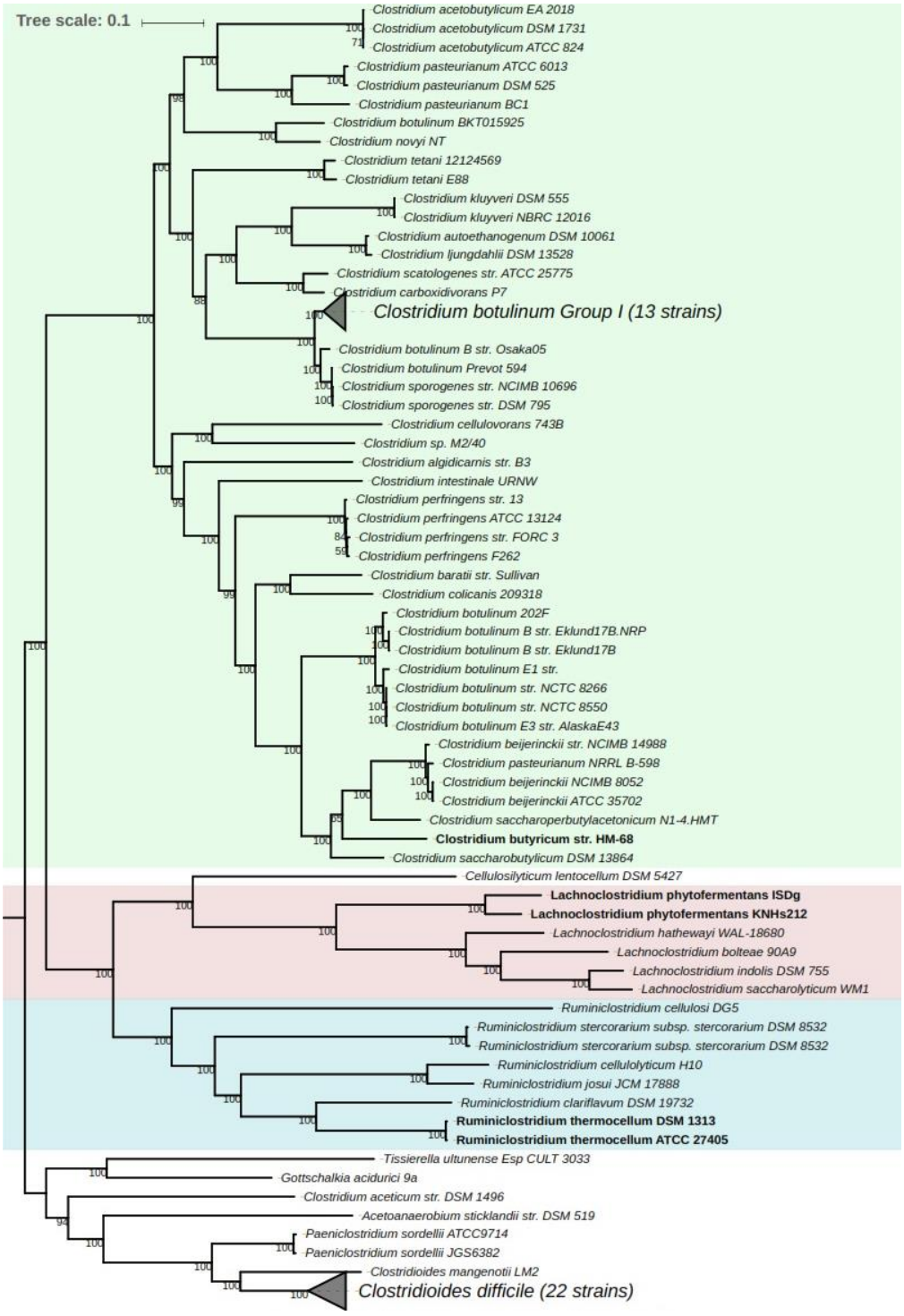


Figure 3.1 Maximum likelihood phylogeny reconstructed using concatenated alignment of 50 conserved housekeeping genes from 101 clostridial strains. Tree visualised using iTOL v3. Bootstrap values (1000 bootstrap replicates) are shown at bipartitions. Bar: branch length. Coloured areas indicate current genera groupings; green: *Clostridium sensu stricto,* pink: *Lachnoclostridium* and blue: *Ruminiclostridium*. For clarity, the *Cl. difficile* and *C. botulinum* Group I nodes have been collapsed and are shown in more detail in Fig. 3.3 A and 3.3 B respectively. Bold type denotes type species.

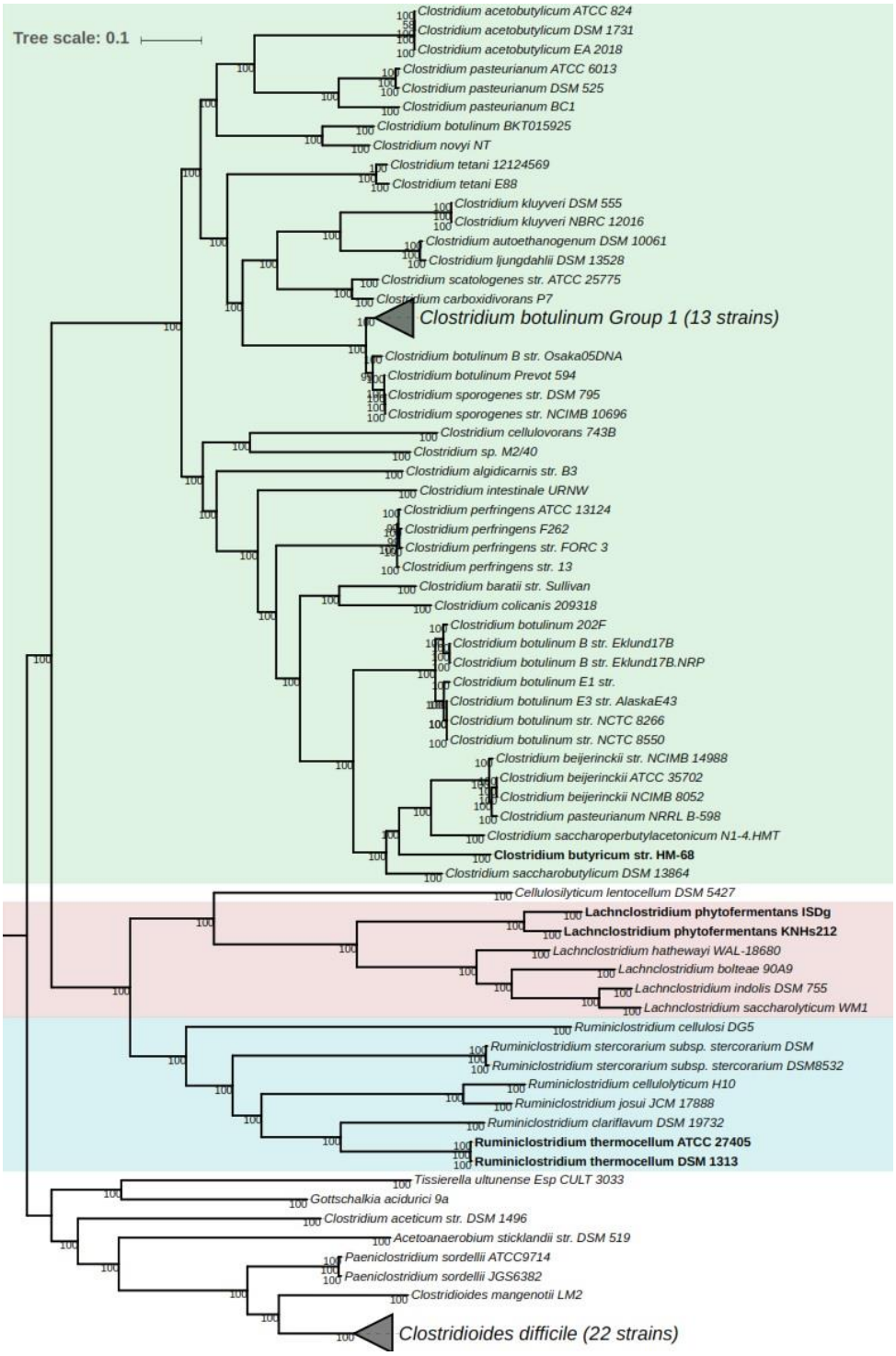


Figure 3.2 Bayesian inference phylogeny reconstructed using concatenated alignment of 50 conserved housekeeping genes from 101 clostridial strains. Tree visualised using iTOL v3. Bayesian posterior probabilities (%) are shown at bipartition. Bar: branch length. Coloured areas indicate current genera groupings; green: *Clostridium sensu stricto,* pink: *Lachnoclostridium* and blue: *Ruminiclostridium*. For clarity, the *C. botulinum* Group I and *Cl. difficile* nodes have been collapsed. Bold type denotes type species.

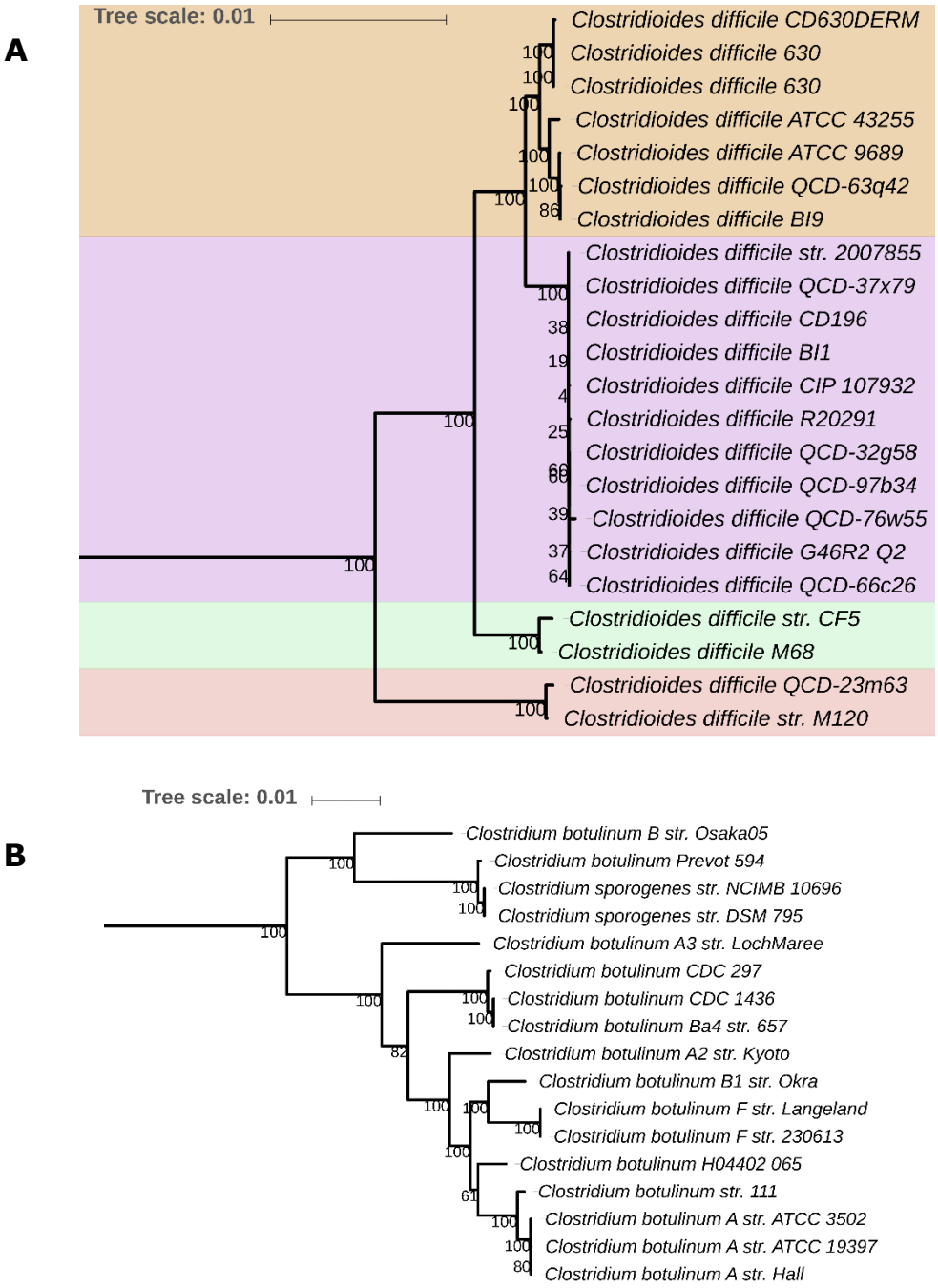
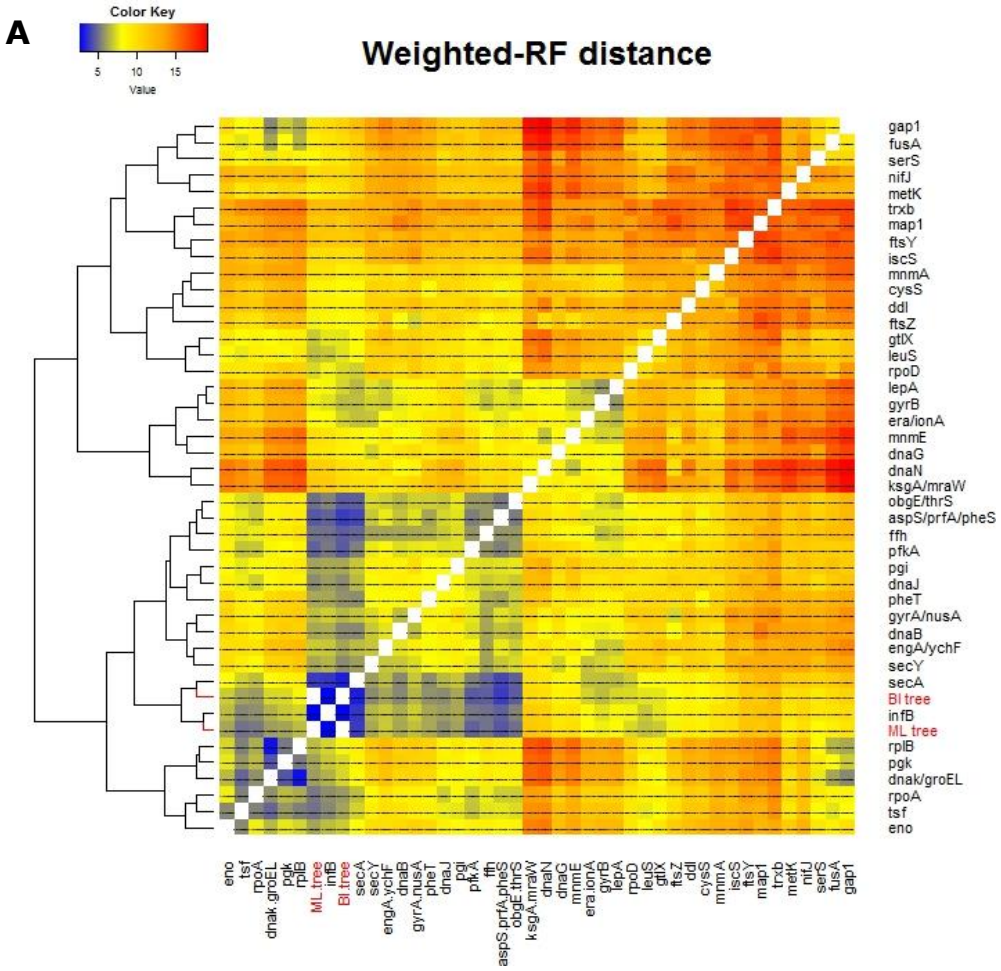
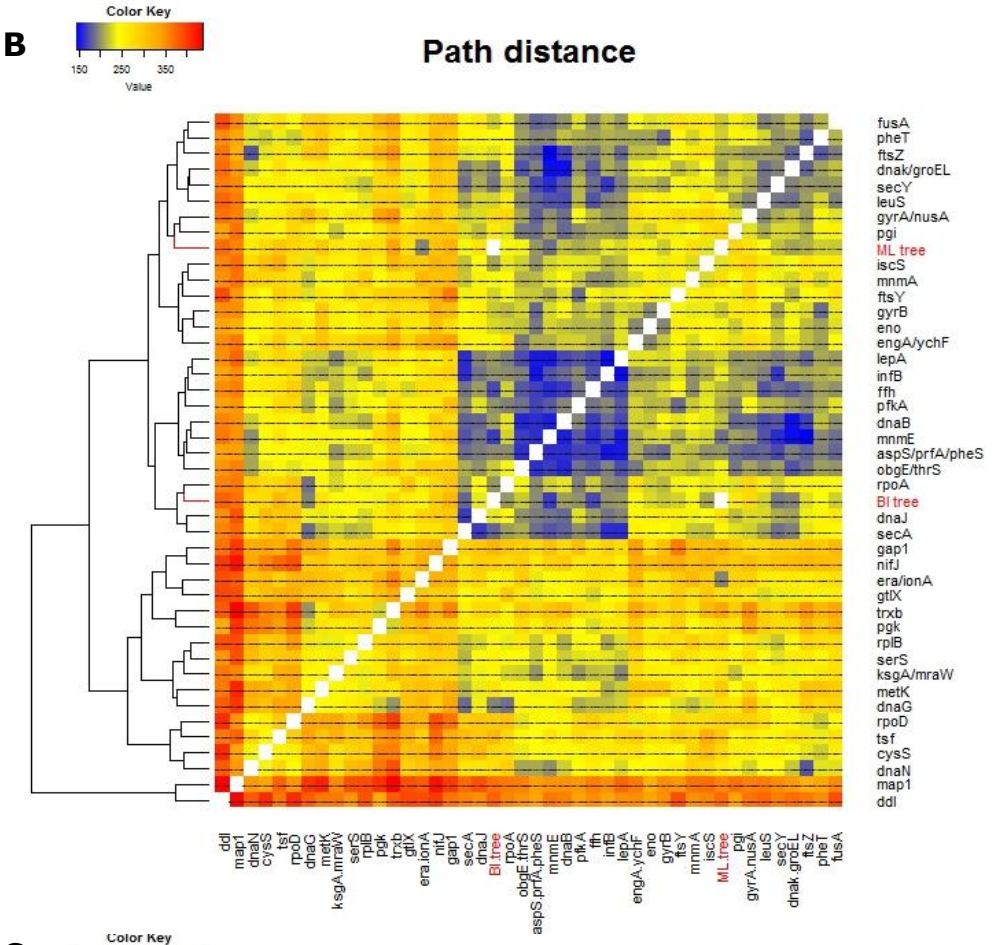
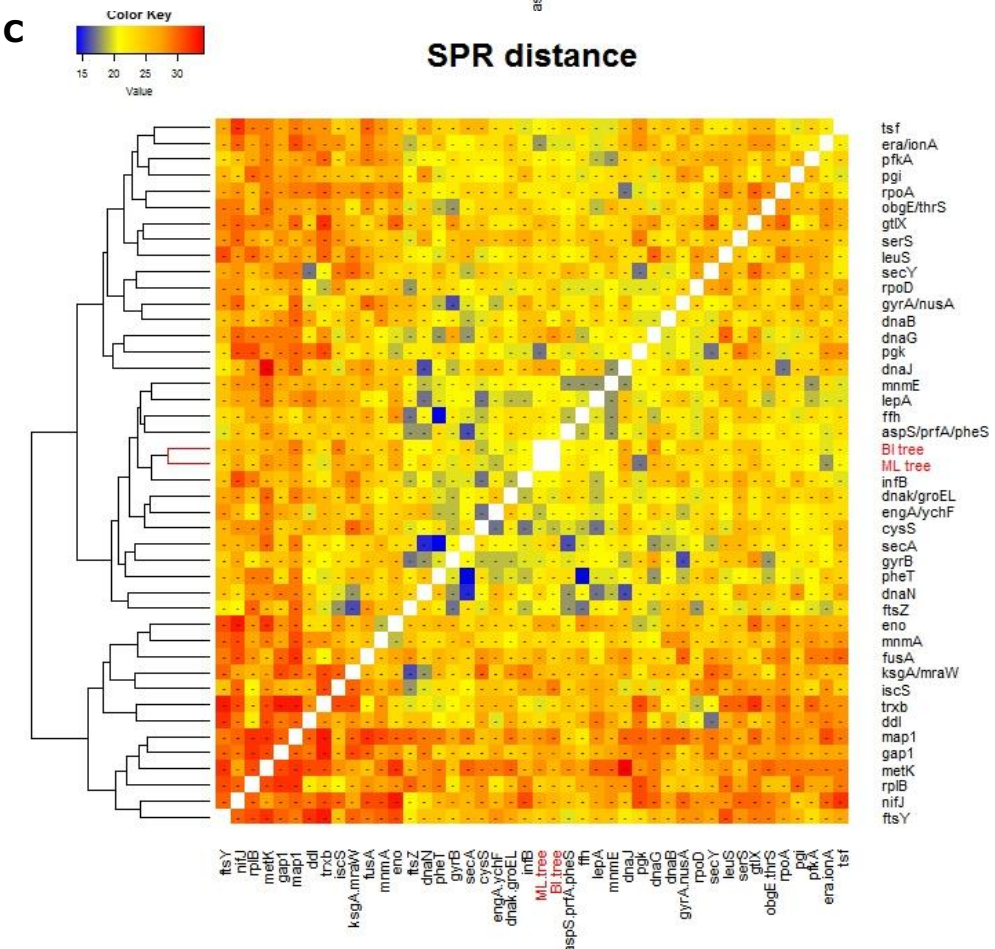


Figure 3.3 Maximum likelihood phylogenies of *Cl. difficile* (A) and *C. botulinum* Group I (proteolytic) (B) clades. Trees visualised using iTOL v3. Bootstrap values (1000 bootstrap replicates) shown at bipartitions. Bar: branch length. Coloured areas denote existing phylogenetic groupings of *C. difficile* proposed by Griffiths *et al.,* (2011); red: Clade 5, green: Clade 4, orange: Clade 1 and purple: Clade 5.


A
Color Key
5
10
15
Value
Weighted-RF distance
gap1
fusA
serS
nifJ
metK
trxb
map1
ftsY
iscS
mnmA
cysS
ddl
ftsZ
gtlX
leuS
rpoD
lepA
gyrB
era/ionA
mnmE
dnaG
dnaN
ksgA/mraW
obgE/thrS
aspS/prfA/pheS
ffh
pfkA
pgi
dnaJ
pheT
gyrA/nusA
dnaB
engA/ychF
secY
secA
Bl tree
infB
ML tree
rplB
pgk
dnak/groEL
rpoA
tsf
eno
eno
tsf
rpoA
dnak.groEL
pgk
rplB
ML.tree
infB
Bl.tree
secA
secY
engA.ychF
dnaB
gyrA.nusA
pheT
dnaJ
pgi
pfkA
ffh
aspS.prfA.pheS
obgE.thrS
ksgA.mraW
dnaN
dnaG
mnmE
era.ionA
gyrB
lepA
rpoD
leuS
gtlX
ftsZ
ddl
cysS
mnmA
iscS
ftsY
map1
trxb
metK
nifJ
serS
fusA
gap1


B
Color Key
150 250 350
Value
Path distance
fusA
pheT
ftsZ
dnak/groEL
secY
leuS
gyrA/nusA
pgi
ML tree
iscS
mnmA
ftsY
gyrB
eno
engA/ychF
lepA
infB
ffh
pfkA
dnaB
mnmE
aspS/prfA/pheS
obgE/thrS
rpoA
Bl tree
dnaJ
secA
gap1
nifJ
era/ionA
gtlX
trxb
pgk
rplB
serS
ksgA/mraW
metK
dnaG
rpoD
tsf
cysS
dnaN
map1
ddl
ddl map1 dnaN cysS tsf rpoD dnaG metK ksgA.mraW serS rplB pgk trxb gtlX era.ionA nifJ gap1 secA dnaJ Bl.tree rpoA obgE.thrS aspS.prfA.pheS mnmE dnaB pfkA ffh infB lepA engA.ychF eno gyrB ftsY mnmA iscS ML.tree pgi gyrA.nusA leuS secY dnak.groEL ftsZ pheT fusA


C
Color Key
15 20 25 30
Value
SPR distance
tsf
era/ionA
pfkA
pgi
rpoA
obgE/thrS
gtlX
serS
leuS
secY
rpoD
gyrA/nusA
dnaB
dnaG
pgk
dnaJ
mnmE
lepA
ffh
aspS/prfA/pheS
Bl tree
ML tree
infB
dnak/groEL
engA/ychF
cysS
secA
gyrB
pheT
dnaN
ftsZ
eno
mnmA
fusA
ksgA/mraW
iscS
trxb
ddl
map1
gap1
metK
rplB
nifJ
ftsY
ftsY nifJ rplB metK gap1 map1 ddl trxb iscS ksgA.mraW fusA mnmA eno ftsZ dnaJ pheT gyrB secA cysS engA.ychF dnak.groEL infB ML.tree Bl.tree aspS.prfA.pheS ffh lepA mnmE dnaJ pgk dnaG dnaB gyrA.nusA rpoD secY leuS serS gtlX obgE.thrS rpoA pgi pfkA era.ionA tsf

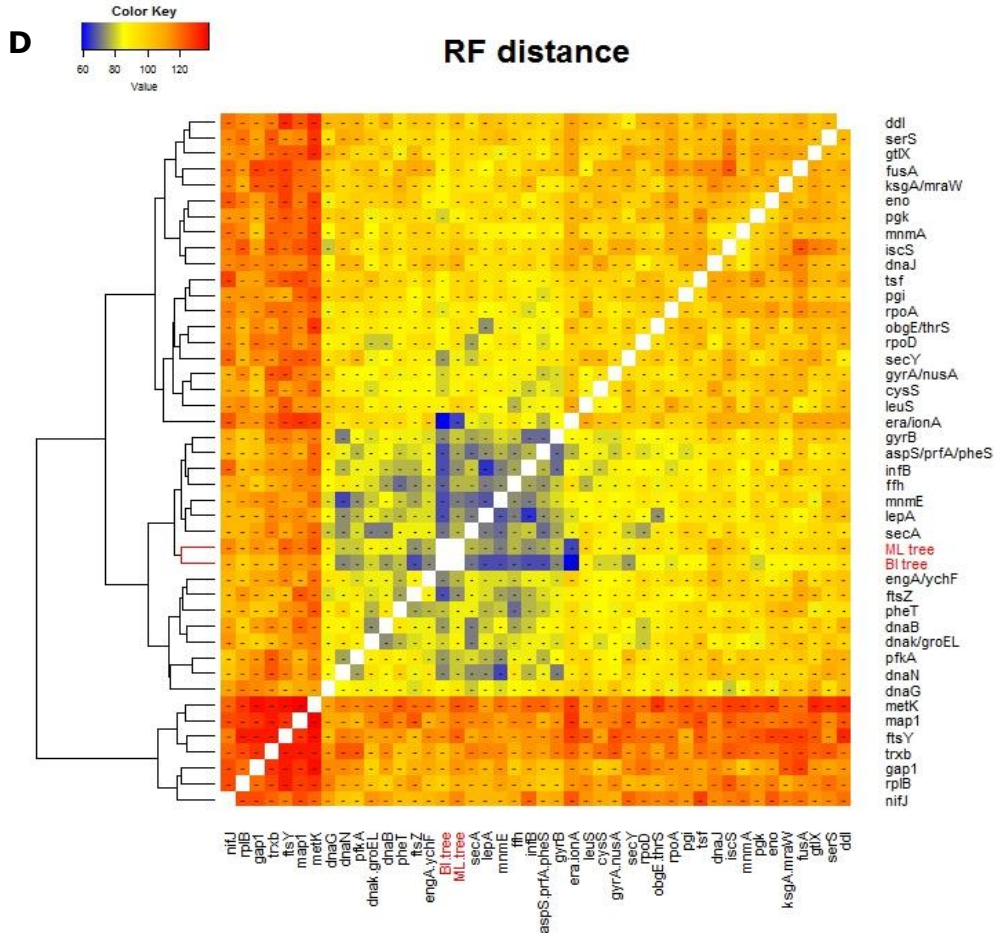


Figure 3.4 Heatmaps showing topological distances between individual partition trees and 50-gene maximum likelihood (ML) and bayesian Inference trees using weighted Robinson–Foulds (A), path difference (B), subtree prune-and-regraft (C) and Robinson–Foulds (D) distance metrics. ML and BI tree metrics in red font. The distance metrics between ML and BI trees were omitted from the heatmaps to improve clarity but gave the lowest difference for all metrics.

3.4.4 Gene discrimination power

Least-square tendency lines were created for all genes and the gene concatenation to infer the discriminative power of the genes. The top 9 discriminating genes (with correlation coefficient R^2 over 0.6) are shown in Figure 3.5, along with the lines for *inf*B and *sec*A genes which showed the closest topology to the concatenated trees. The phylogenetic distances of the concatenated sequence least correlated with *map, gtl*X and *isc*S genes with R^2 coefficients below 0.4 (data not shown). The *dna*N and *trx*B genes

were twice as discriminatory as the concatenated gene sequence (Figure 5), while *inf*B and *sec*A genes were comparable to that of the concatenated sequence (Figure 3.5). The *gro*EL, *pgi* and *rpl*B genes were the least discriminatory, having 65.8, 63.8 and 51.1% of the discriminative power of the gene concatenation (data not shown).

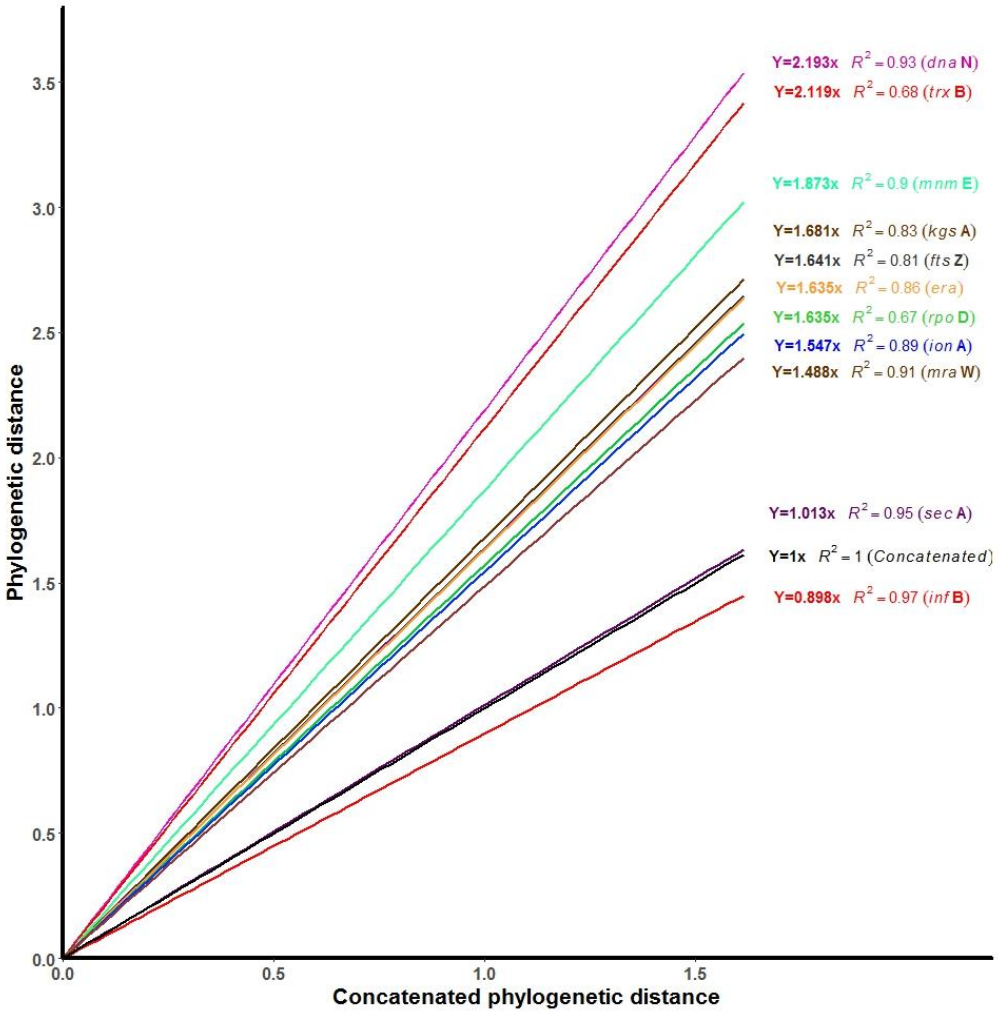


Figure 3.5 Least square tendency lines generated by regressing pairwise phylogenetic distances between each gene and the 50 gene concatenation. Lines have been plotted through the intercept for clarity. The slope and R^2 coefficients are given for the 11 genes shown, along with the 1 v 1 line of the concatenated line.

3.4.5 InfB based phylogenies

The *inf*B gene was selected to further explore clostridial phylogeny using additional sequences from NCBI's larger whole-genome shotgun contigs database. Using BLASTN, putative *inf*B sequences for 238 *Clostridium* strains were retrieved from the BLASTN hit results. These were processed and aligned as described previously, and an ML phylogeny was produced (Figures 3.6 A & B.) The estimates of species ancestry in the *infB* tree are similar to the *Clostridium sensu stricto* clade in the multi-gene phylogeny, the main exception being that in the *infB* phylogeny, the *C.novyi/acetobutylicum* subclade is more closely related to the large *C. botulinum* Group II-containing subclade (Figure 3.6 A) than the *C. botulinum* Group I/*tetani* subclade (Figure 3.6 B). One clinically important taxonomic conflict is within the *C. botulinum* Group III/ *C. novyi sensu lato* clade, shown in detail in Figure 3.7. There are 5 distinct lineages containing *C. botulinum* Types C, D, C/D, *C. novyi* types A and B, *C. haemolyticum* and *C. massiliodoelmoense,* although these lineages do not correlate with existing species designations.


A
Clostridium bornimense replicon M2/40 rep1
Clostridium cellulovorans 743B
Clostridium grantii DSM 8605
Clostridium homopropionicum (2)
Clostridium colicanis DSM 13634 CLCO contig000002
Clostridium thermopalmarium DSM 5974 CPAL contig000012
Clostridium botulinum group III/ novyi sensu lato (26)
Clostridium acidisoli DSM 12555
Clostridium akagii DSM 12554 BR66DRAFT scaffold00008.8 C
Clostridium pasteurianium (3)
Clostridium pasteurianum BC1
Clostridium arbusti SL206
Clostridium sp. DMHC 10 contig 15
Clostridium sp. CT7 contig.1
Clostridium pasteurianum strain GL11 contig.1
Clostridium acetobutylicum (4)
Clostridium roseum strain DSM 6424 CLROS contig000017
Clostridium roseum strain DSM 7320 CROST contig000076
Clostridium felsineum DSM 794 CLFE contig000002
Clostridium aurantibutyricum strain DSM 793 CLAUR contig000063
Clostridiales bacterium oral taxon 876 str. F0540 Scaffold37
Clostridium collagenovorans DSM 3089
Clostridium estertheticum subsp. estertheticum strain DSM 8809
Clostridium sulfidigenes strain 113A c23
Clostridium tunisiense TJ C661contig6
Clostridium argentinense CDC 2741 U732.Contig247
Clostridium ihumii AP5
Clostridium sp. ATCC 25772 U20725 10
Clostridium senegalense JC122 strain type strain JC122
Clostridium oryzae strain DSM 28571 CLORY contig000019
Clostridium sp. N3C
Clostridium polynesiense strain MS1
Desnuesiella massiliensis strain mt10
Clostridium amazonitimonense
Clostridium algidicarnis (2)
Clostridium cadaveris (3)
Clostridium hydrogeniformans DSM 21757 BR69DRAFT scaffold00001.1
Clostridium frigidicarnis strain DSM 12271
Clostridium intestinale (2)
Clostridium fallax strain DSM 2631
Clostridium cavendishii DSM 21758
Clostridium mediterraneense strain Marseille-P2434
Clostridium sp. ND2
Clostridium perfringens (14)
Clostridium ventrilculi (3)
Clostridium sp. CL-2
Clostridium disporicum strain 2789STDY5834856
Clostridium celatum DSM 1785 Scfld538
Clostridium disporicum strain 2789STDY5608827
Clostridium disporicum strain 2789STDY5834855
Clostridium saudiense strain JCC
Clostridium gasigenes strain DSM 12272
Clostridium septicum strain P1044
Clostridium chauvoei (3)
Clostridium nigeriense strain Marseille-P2414
Clostridium sp. C8 contig00042
Clostridium tertium (2)
Clostridium sp. DSM 8431
Clostridium vincentii strain DSM 10228 CLVI contig000015
Clostridium paraputrificum (4)
Clostridium thermobutyricum DSM 4928 CLTH contig000006
Clostridium sp. Ade.TY contig22
Clostridium baratii (6)
Clostridium botulinum group II (5)
Clostridium sacchrobutylicum (4)
Clostridium neonatale (2)
Clostridium butyricum (10)
Clostridium puniceum strain DSM 2619 CLPUN contig000223
Clostridium sp. BL-8 CLOBL contig000047
Clostridium sp. DL-VIII chromosome
Clostridium chromiireducens strain DSM 23318 CLCHR contig000007
Clostridium sacchroperbutylacetonicum (2)
Clostridium beijerinckii (10)
Clostridium beijerinckii strain NRRL B-591 CLBKI contig000172
Clostridium diolis strain NJP7 scaffold3
Clostridium diolis strain WST Scaffold13 1
Clostridium beijerinckii strain Br21 CBEIBR21 SCAFF02
Clostridium beijerinckii strain NCIMB 14988
Clostridium beijerinckii strain BAS/B2 CLBEJ contig000013
Tree scale: 0.1

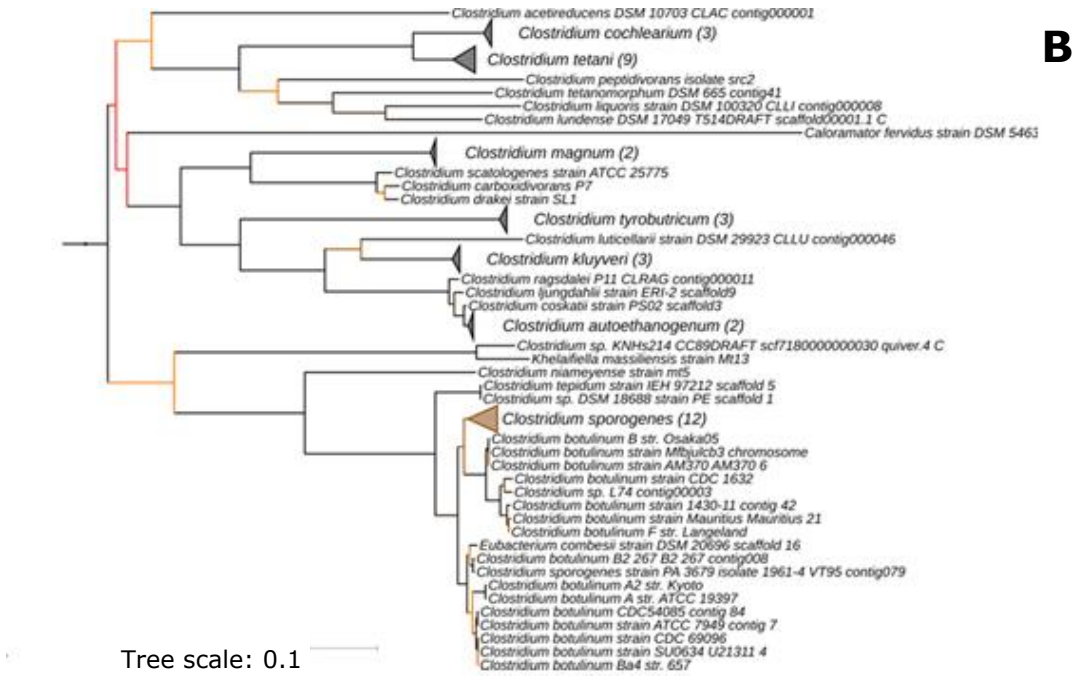


Figure 3.6 *Inf*B gene phylogeny inferred by maximum-likelihood of 238 *Clostridium sensu stricto* strains falls into two distinct clades shown in A and B. For clarity, monophyletic clades contain taxa correlating to one species designation have been collapsed. Coloured branches indicate bipartition bootstrap support (1000 replicates), brown <95%, orange <75% and red <50% support.

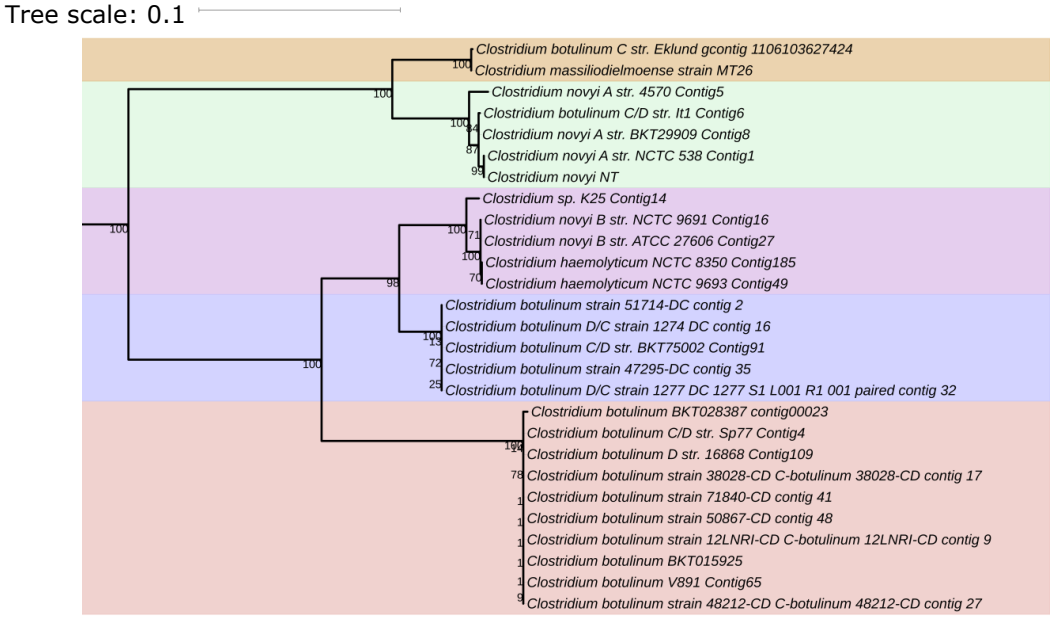


Figure 3.7 The *inf*B gene tree of the *Clostridium novyi sensu lato clade*. Trees inferred by ML and visualised using iTOL v3. Bootstrap values (1000 bootstrap replicates) shown at bipartitions. Bar: branch length. Coloured areas indicate distinct lineages within the clade.

3.5 DISCUSSION

Class Clostridia has been subject to recent taxonomic reshuffling (see Gupta and Gao (2009) and Yutin and Galperin (2013)) due to the increasing availability of nucleotide coding sequences as well as improved bioinformatics tools, enabling better phylogenetic reconstruction. As a result, many former *Clostridium* species have been reassigned to alternative or new genera, although many species have kept their original names. Antiquated nomenclature can prove problematic for researchers working with clostridial species. This warranted a fresh look at clostridial phylogeny and the establishment of a simple, robust method for generating accurate phylogenies universally. Freely-available phylogenetic software was utilised to develop a simple, reproducible bioinformatics approach for process existing and newly-available clostridial genomic material.

Homologous nucleotide coding sequences were retrieved, inspected and analysed for 50 essential housekeeping genes from 101 clostridial taxa. Phylogenetic trees were constructed for each gene, and for a 50-gene concatenated sequence which was hypothesised to give the closest approximation of the true phylogeny. The concatenated phylogenies emulate the recent clostridial species reclassification designation suggested by Yutin and Galperin (2013) while providing a high level of discriminative power for individual strains within 13 different species. Least-square tendency line analysis and tree comparison algorithms were used to identify genes which produced similar phylogenies to the concatenated sequences and could be used as reliable markers for clostridial species and subspecies differentiation. Compared to the ML concatenation, the *inf*B and *sec*A genes correlated well with the genetic distances (R^2= 0.97 and 0.95, respectively) and were almost equally discriminating. Furthermore, *inf*B and *sec*A individual gene trees showed consistently low differences in tree topology based on various distance metrics. The *inf*B nucleotide coding sequences were used to produce detailed phylogenies for genus *Clostridium sensu stricto* species. The *inf*B

revealed interesting strain groupings for some pathogenic strains, which are likely to be of clinical significance.

3.5.1 Species distinctions

Multigene phylogenies identified a distinct clade of *Clostridium* spp. i.e. *Clostridium sensu stricto,* which then splits into two distinct clades. This separation was also observed in the *inf*B phylogeny with 90 validly-named *Clostridium* species. (Figure 3.6); with only slight discrepancies in ancestry. The analysis conducted in this study clearly placed the organism *Clostridium aceticum* outside the *Clostridium sensu stricto* clade and other well-defined clades (Figures 3.1 and 3.2). *C. aceticum* is listed as a true *Clostridium* on key resources such as NCBI Taxonomy Database, UniProt (The UniProt Consortium 2011)], Kyoto Encyclopaedia of Genes and Genomes [KEGG, http://www.kegg.jp/kegg/genome.html] and List of Prokaryotic Names with Standing in Nomenclature (Euzéby, 1997). This segregation of *C. aceticum* from the *Clostridium* genus been reported previously (Ye *et al*., 2004; Cao, Liu and Dong, 2003; Poehlein *et al*., 2015; Pikuta *et al*., 2006), and the species is not included within the genus in Taxonomic Outlines for Volumes 3 of Bergey's Manual of Systematic Bacteriology (Second Edition) (Ludwig, Schleifer and Whitman, 2009). Continued use of this nomenclature may be misleading.

The grouping of toxigenic *C. botulinum* str. Prevot 594 and C. *botulinum* str. Osaka05 with two non-pathogenic *C. sporogenes* strains is consistent with other recent studies, which found that some toxigenic *C. botulinum* Group I strains, including Prevot 594 and Osaka05, were more closely related to *C. sporogenes* strains based on whole genome or orthologous gene analysis (Smith *et al.*, 2015; Weigand *et al.*, 2015). These *C. sporogenes* bacteria have seemingly acquired the neurotoxin genes *via* horizontal transfer of plasmid DNA. It was also demonstrated that toxigenic strains can lose their toxicity (Weigand *et al.*, 2015). In the *inf*B phylogeny the apparent misclassification of *C. botulinum* Group III, *C. novyi* and *C. haemolyticum* strains is observed (Figure 3.7). This phenomenon was

discussed in a recent study which characterised this complex subgroup using genome and plasmid sequences (Skarin and Segerman, 2014). The 4 larger subgroups are in exact agreement with the lineage distinctions identified by Skarin and Segerman (2014), and the 5th subgroup identified by our *inf*B phylogeny (containing *C. botulinum* C str. Eklund and *C. massiliodoelmoense*) could represent a new distinct lineage within this group. It is likely that these subgroups are also capable of phage-driven plasmid and toxin gene transfer between other members of the *C. novyi sensu lato* group. The multigene phylogeny also clustered the 22 strains of *Clostridioides difficile* into 4 main clades (Figure 3.3 A). These groupings are consistent with both toxin-gene, MLST and genome-based clustering patterns found in other studies (Dingle *et al*., 2011; Donskey *et al*., 2014; Persson, Torpdahl and Olsen, 2008; Griffiths *et al*., 2010; Chowdhury *et al*., 2016), and allow differentiation between strains with different levels of virulence and infectiousness.

The species *[C.] ultunense* was not assigned to a genus by Yutin and Galperin (2013). Although initially classified as a *Clostridium* sp. (Schnurer, Schink and Svensson, 1996), its nomenclature has since been changed to reflect that it is not a true *Clostridium* representative, as supported by the analysis presented here. A BLASTN search of *[C.] ultunense* 16S rDNA revealed that the only fully-named related species is *Keratinibaculum paraultunense*. Despite this similarity, a number of phenotypic differences have been identified between the two species, including growth parameters and substrate utilisation, plus the gelatin hydrolysis, lack of flagella and motility and the lower G + C% demonstrated in *K. paraultunense* compared to *[C.] ultunense* (Huang *et al.*, 2013). *[C.] ultunense* has been found to have a very similar phenotype to *Tepidimicrobium* spp. (Huang *et al.*, 2013), although a strong 16S rRNA sequence similarity was not revealed by a BLAST search.

3.5.2 Species recombination and misnamed strains

Phylogenetic analysis showed that the important industrial species *C. autoethanogenum* and *C. ljungdahlii* are very closely related genetically, showing 98.8% sequence identity across the 69,578-position alignment. However, other studies have shown distinct phenotypic and metabolic differences between the two taxa (Tanner, Miller and Yang, 1993; Cotter, Chinn and Grunden, 2009). Brown *et al.,* (2014) revealed various genetic distinctions between the two. The *inf*B phylogeny supports the species distinction. Other species such as *C. colicanis*/*C. thermopalmarium* and *C. roseum/C. felsineum/C. aurantibutyricum* show small or no differences evolutionary distances based on tree branch lengths of this gene tree. Further investigation is warranted to explore whether these species truly warrant division.

Furthermore, numerous potentially misnamed strains were identified. *C. pasteurianum* str. NRRL B-598 clustered within the *C. beijerinckii* clade. This genome had a considerably larger genome size (6.19 Mb) compared to that of type strain *C. pasteurianum* str. ATCC 6013 (4.35 Mb) and str. BC1 (4.99 Mb) (Sedlar *et al*., 2015). The NRRL B-598 genome is more similar in size to *C. beijerinckii* strains; ATCC 35702 (5.99 Mb), NCIMB 14988 (6.48 Mb) and NCIMB 8052 (6.00 Mb). Additionally, NRRL B-598 is unable to utilise glycerol unlike other *C. pasteurianum* strains (Sedlar *et al*., 2014), while the Spo0A master regulation of sporulation gene only differs by one amino acid from that of *C. beijerinckii* NCIMB 8052 (Sedlar *et al*., 2015). Other recent studies have noted similar physiology to *C. beijerinckii* (Kolek *et al*., 2015), and combined with the strong genetic relatedness shown in this study, these results conceive that *C. pasteurianum* str. NRRL B-598 is an atypical strain of *C. beijerinckii.*

3.6 CONCLUSIONS

This research demonstrates the high discriminatory power of the methods utilised in this study, and that multiple housekeeping-gene based approaches can be utilised for accurate reconstruction of phylogenies at genus, species and strain level. The phylogenetic analysis of new genome sequences, using the methodology detailed in this paper, should allow a better taxonomic assignment than is achieved using 16S rRNA gene sequences alone, when used in combination with phenotypic characteristics. This study supports the recent species reassignment by Yutin and Galperin (2013), while also identifying the misplacement of *C. aceticum*. This study also highlights the issues regarding strain identification based on toxin genes or phenotypic characteristics, which has led to the incorrect taxonomic assignment of numerous strains, including some pathogenic strains. Comparison of tree topologies and gene discrimination power revealed that the *infB* gene was a more suitable candidate for single-gene phylogenetic inference. Using this gene, it was possible to correctly distinguish between clinically and phylogenetically separate strains and pathotypes of pathogenic species. This represents an equally simple, but more accurate method for identifying and classifying clostridial strains compared to the 16S rRNA gene, while being more cost-effective than whole-genome phylogenetic analysis. All bioinformatics work was performed on Windows using freely-available software, enhancing the usefulness and applicability of these methods to other researchers wishing to perform similar phylogenetic analysis.

CHAPTER 4: SURVIVAL OF PATHOGENIC CLOSTRIDIA IN WATER AMENDED SOIL

4.1 ABSTRACT

Survival of clostridial pathogens following introduction into the soil is poorly documented. The anaerobic, spore-forming nature of these organisms is likely to promote different behaviour in the soil in response to abiotic and biotic factors compared with other well-studied enteric pathogens. Microcosm experiments were established and real-time PCR (RT-qPCR) was used to investigate survival of *Clostridioides difficile*, *Clostridium perfringens* and *Clostridium sporogenes* (a surrogate for proteolytic *C. botulinum*) in response to soil water status. Microcosms were maintained at 25% water-holding capacity (WHC), 75% WHC or were flooded; each treatment represented environmentally realistic values. All species persisted for >84 days under all treatments, despite a rapid decline in the first week. Survival was significantly greater for all species in soils adjusted to 25% WHC than in wetter soils. First stage decay coefficients of 0.11 – 0.67 (gene copies) day^{-1} were derived from bioexponential decay models, with the decay rates significantly lower in drier soils ($p = 0.026$). Pathogens in the drier soils transitioned earlier (5-6 days) to a slower stage of decay (<0.01 – 0.05) day^{-1}) than those in wetter soils (days 9-12) ($p = 0.004$). This second stage likely represents pathogens in the spore state. Earlier transition into spores by *Clostridium* spp. maintained in the driest soils is likely to maintain pathogen abundance, which may indicate an increased health risk by direct contact-mediated infection from contaminated dry soil or soil dust.

Keywords; Bacterial survival; die-off; *Clostridioides difficile; Clostridium perfringens; Clostridium botulinum*

4.2 Introduction

Soils are an important reservoir for a range of pathogenic microorganisms. While survival and persistence of some zoonotic pathogens such as *Escherichia coli*, *Salmonella*, *Listeria monocytogenes* and *Cryptosporidium parvum* are well documented (Jamieson *et al*., 2002), the fate of clostridial pathogens is poorly understood. Enteric pathogens including *Clostridioides difficile* and *Clostridium perfringens* can be introduced to agricultural soils through faeces, manure and feedlot runoff water. Other organic soil-amendments (OSA) such as compost, slurry and anaerobic digestates can harbour (and potentially facilitate proliferation of) both enteric and non-enteric clostridial pathogens like *Cl. difficile, C. perfringens and Clostridium botulinum* (Böhnel and Lube, 2008; Avery *et al*., 2012; Neuhaus *et al*., 2015; Xu *et al*., 2016; Gessler and Böhnel, 2006), which can also contaminate agricultural environments. While attempts have been made to undertake risk assessments regarding Clostridia in OSAs (Böhnel, Briese and Gessler, 2002), a key problem is the lack of background prevalence and survival data in agricultural soils. If these soils act as a major reservoir for these pathogens, they can act as a source for infection and reinfection of livestock, contamination of produce (Girardin *et al*., 2005; Stefanis *et al*., 2014) and water courses and direct contact-mediated infection of humans (soil ingestion, dust inhalation and wound infection exposure pathways).

Once introduced to the soil, the persistence or die-off of the pathogen is affected by a multiplicity of factors as described in Chapter 2 and reviewed by other authors (Jamieson *et al*., 2002; Reddy, Khaleel and Overcash, 1981; Crane and Moore, 1986). Wet or saturated soils may create oxygen-deprived, anoxic environments that could facilitate reduced die-off or even growth of anaerobic pathogens (Tiedje *et al*., 1984). While the long-term persistence of clostridial pathogens after introduction to the soil has been demonstrated (Gessler and Böhnel, 2006; Girardin *et al*., 2005; Sandler *et al*., 1993), evidence of proliferation is not documented. The saccharolytic,

polysaccharolytic, proteolytic and nitrogen-fixing capabilities of many clostridial species may allow pathogens to naturalise under certain soil conditions. One long-term field study demonstrated the recovery of *C. perfringens* populations after a cold winter period (Brochier *et al*., 2012), indicating the capability for growth in the soil, rather than merely persisting. Indeed, Clostridia have been found to be the dominant taxa in anaerobic environments such as rice paddy soil (Weber, Stubner and Conrad, 2001; Liu *et al*., 2009; Liesack, Schnell and Revsbech, 2000), wet tropical soils (Pett-Ridge and Firestone, 2005) and sediments (Williamson, Rocke and Aiken, 1999; Davies *et al*., 1995), and it follows that Clostridia may be prevalent in regularly or permanently saturated soils. For example, the use of biological soil disinfestation, which creates anaerobic conditions to kill crop pathogens, was found to create soils dominated by Clostridia (Mowlick *et al*., 2012; Rood and Cole, 1991). To date, persistence of pathogenic Clostridia in temperate waterlogged soils has not yet been researched. However, despite the seemingly clear link between wet soils and increased clostridial growth, Clostridia species are frequently isolated from both wet and dry soils, and are often reported as an abundant taxa in many soil bacterial diversity studies (Dorr de Quadros *et al*., 2012; Russo *et al*., 2012; Janssen, 2006). The ability of Clostridia to sporulate is likely to promote their persistence in dry soils and soil dust, explaining their perceived prevalence in soils globally. Environmentally-resistant clostridial spores lay dormant in the soil, potentially for hundreds of years (Setlow, 2007). The spore coat also protects clostridial spores from protozoan predation (Klobutcher, Ragkousi and Setlow, 2006) and desiccation in periods of drought (Setlow, 2007). This could result in comparatively lower rates of die-off (longer persistence) than for non-spore forming bacteria.

Determining biologically-meaningful die-off rates for clostridial pathogens in agricultural soils can directly help guide policy makers and landowners in managing risk associated with agricultural practices. Once death-rates have been established, the residence time for individual pathogens in the

soil can be determined for a given range of soil and environmental parameters and the efficacy of management interventions can be modelled.

The objective of this study was to investigate the impact of soil water content on the survival and persistence of the important clostridial pathogens *Clostridioides difficile, C. perfringens* and *C. sporogenes* in a British, agricultural soil. It is hypothesised that i) pathogen numbers decrease over time in a biphasic manner and ii) the rate of decay is decreased with higher water content.

4.3 EXPERIMENTAL PROCEDURES

4.3.1 Bacterial strains

A cocktail of three clostridial species was used to spike soil microcosms. These were the cytotoxin-positive *Clostridioides difficile* NCTC 13307 (Public Heath England, UK), epsilon toxin–negative *Clostridium perfringens* NCTC 13170 and *Clostridium sporogenes* ATCC 19404 (Sigma-Aldrich, UK). *C. sporogenes* was selected as a surrogate for Group I (proteolytic) *Clostridium botulinum* due its genetic and phenotypic similarity to the neurotoxic bacterium, but lack of pathogenicity. Cultures were inoculated separately into 300 ml of sterile thioglycolate broth with resazurine (Sigma-Aldrich, UK), and incubated at 37°C for 50 hrs in anaerobic jars containing anaerobic atmosphere-generating packs (Oxoid, UK). The initial inoculation was performed in an anaerobic cabinet (Don Whitley Scientific, UK) to ensure anaerobic conditions were maintained throughout, thus minimising oxidative stress to cells (Edwards, Suárez and McBride, 2013).

Cells were then collected by centrifugation (10 minutes; 5,000 × g; 10°C; Allegra X-22D centrifuge) in sterile 50 ml Falcon tubes, then resuspended in 20 ml of 0.01 M phosphate buffered saline (PBS) (Sigma-Aldrich, UK). This was repeated twice, and cells were finally eluted in 5 ml of sterile deionised water (SDIW). The pellet was gently vortexed to suspend and pooled together with cleaned cell suspensions from the other two clostridial

pathogens. This pooled cell suspension was inverted 5 times to mix and 5 ml was decanted into a sterile 50 ml Falcon tube. This cell suspension was then diluted 1:10 with SDIW and used immediately to spike soil microcosms.

4.3.2 Soil microcosm design

Soil (pH 5.65) was collected from the upper 10 cm of a James Hutton Institute agricultural plot at Balruddery Farm, Dundee (56°28'N, 3°7'W). The arable soil, defined as a Cambisol (FAO Unesco classification), was typical of the region. The soil was a free-draining, sandy loam, identified in 2013 as belonging to the Balrownie Association, series Buchanyhill, and was in between potato cultivations (vegetation classification AR4A). After collection, the soil was stored at 4°C for two weeks prior to microcosm construction.

A 3 kg subsample of the soil was dried at 30°C for 48 hrs and sieved (2 mm mesh size). The water holding capacity (WHC) was calculated using 100 g of dried, sieved soil which was placed in a metal ring with a filter-paper at the base. The soil was saturated with DI water, sealed with foil (to prevent evaporation losses), and left for 12 hrs to drain. The soil was then reweighed (deducting weight of metal ring and filter paper) before oven drying at 120°C for 24 hrs before reweighing. This was performed in triplicate, and the WHC of the soil was derived using the formula:

$$WHC\left(\frac{ml}{100}g\ dwt\ soil\right) = \frac{Mass(g) of\ saturated\ soil - Mass\ (g)\ of\ oven\ dried\ soil}{Mass\ (g) of\ oven\ dried\ soil} \quad (1)$$

Microcosms were constructed by loading 25 g of dried, sieved soil to sterile, 60 ml polypropylene flasks. Two trials were conducted. In a preliminary experiment examining pathogen die-off within the first seven days, twenty-four soil microcosms were amended to two water treatments; 20% WHC or flooded, by addition of artificial rainwater (see Table 4.1) spiked

with *C. sporogenes* ATCC 19404 cell solution. The main experiment evaluated pathogen die-off over eight-four days. Sixty-six microcosms were constructed and assigned to one of three water treatments; 25% WHC, 75% WHC or flooded. This was achieved by gentle addition by pipette of 3.6 ml, 10.86 ml, or 17 ml of sterile artificial rainwater respectively, spiked with a combination of *Cl. difficile* NCTC 13307, *C. perfringens* NCTC 13170 and *C. sporogenes* ATCC 19404. Three replicated microcosms from each water treatment were not spiked, in order to investigate the regrowth of indigenous clostridial pathogens with the soil. An additional three microcosms were not assigned to any treatment and were used to calculate the abundance of indigenous Clostridia already in the soil. After spiking, flasks were sealed, weighed and stored at room temperature in the dark. Every three days, soils were aerated in a sterile Category II Biological Safety cabinet to prevent establishment of anaerobic conditions in the headspace above the soil. Microcosms were also reweighed, and if necessary, amended to their original weight by addition of sterile rainwater to maintain the correct experimental WHC to adjust for any moisture evaporation into the headspace.

4.3.3 Soil sampling

Microcosms were destructively sampled on days 0 (immediately after spiking), 4 and 7 for the preliminary experiment, and days 0, 7, 14, 21, 42 and 84 for the main study. At the time of sampling, microcosms were randomly selected from each treatment in triplicate. The microcosms under 20%, 25% and 75% WHC treatments were amended to the same water concentration as the flooded microcosms, by addition of sterile artificial rainwater. Microcosms were homogenised with a sterile polypropylene stirrer and sampled using sterile polypropylene drinking straws to remove 1-2 g soil core. Soil samples were placed in sterile 2 ml cryogenic tubes, immediately snap frozen in liquid nitrogen and stored at -20°C until DNA extraction. All sampling was performed in a sterile Category II Biological Safety cabinet.

Table 4.1 Concentrations of chemical components in rainwater and artificial rainwater

Component	Target range in artificial rain µM l^{-1}	Chemical	µM l^{-1} in artificial rain	g l^{-1} required for 250ml of 1000 × stock
Ca	5-10	$CaCl_2 \bullet 2H_2O$	7.5	0.28
Cl	101-280	KCl	4	0.07
H^+	10-25	KH_2PO_4	3.5	0.12
K	2-10	$MgCl \bullet 6H_2O$	20	1.02
Mg	10-31	NaCl	40	0.66
Na	83-277	Na_2SO_4	17.5	0.82
NH_4-N	5-26	NH_4NO_3	20	0.40
NO_3-N	6-25			
P	2-5			
S	11-31			

4.3.4 DNA extraction and preparation

Genomic DNA (gDNA) was extracted from soil using the PowerSoil® DNA 96-well Isolation Kit (Mobio, Inc. Carlsbad, CA, USA). On the day of extraction, soils were thawed on ice and 250 mg of soil was added aseptically to an individual well of the bead-beading plate. DNA was extracted following the manufacturer's instructions, except with the inclusion of 2 freeze/thaw cycles after the addition of the PowerSoil® Bead Solution (stored at -70°C until frozen then immediately thawed in a 65°C water bath), followed by heating to 70°C for 15 minutes prior to bead beating. These additional steps were recommended by the manufacturer to help improve the lysing efficiency from bacterial spores. After extraction, gDNA yield was quantified using a Qubit® 2.0 Fluorometer (Invitrogen, Paisley, UK) with the Qubit® Broad Range assay. The quantity and purity

of gDNA, based on absorbance ratio at both 260/280 and 230/260 nm, was also determined by UV spectrophotometry using a NanoDrop instrument (Thermo Fisher Scientific, Hemel Hempstead, UK). For both techniques, 1 µl of gDNA was used.

4.3.5 Real-time quantitative PCR analysis

The survival of *Cl. difficile, C. perfringens* and *C. sporogenes* was determined by real-time quantitative PCR (RT-qPCR). Commercially available kits (Genesig®; PrimerDesign Ltd., Southampton, UK) were used to quantify target gene copy-numbers in extracted gDNA. FAM-labelled primer/probe mixes were used to target toxin A *(toxA*)-producing *Cl. difficile*, alpha-toxin (*cpa)*-producing *C. perfringens* and 16S ribosomal RNA genes in *C. sporogenes*. Four µl of the supplied internal extraction control DNA (*Rattus rattus)* was spiked into the C1 lysis buffer during the DNA extraction protocol detailed previously. This DNA was amplified using a VIC-labelled primer/probe set and was used to calculate the efficiency of the DNA extraction process and any potential PCR-inhibition. The 20 µl qPCR reaction contained; 15 µl reaction mix consisting of 10 µl PrecisionPLUS™ 2 X qPCR Mastermix, 1 µL FAM-labelled target-specific primer/probe mix, 1 µl of VIC-labelled internal control primer/probe mix and 3 µl of RNAse/DNAse free water (Sigma-Aldrich, UK), and 5 µl of gDNA sample template, normalised to 2 ng ul^{-1}. The positive control DNA provided was used to generate standard curves for each assay (10 - 10^5 gene copies), in triplicate, along with sterile molecular water no-template controls (NTC). Two technical replicates (for qPCR quality assurance purposes) were performed for each microcosm DNA sample, on separate Hard-Shell® white 96-well PCR plates (Bio-Rad, UK), and sealed with Microseal B adhesive seals (Bio-Rad, UK). The RT-qPCR was performed on a Bio-Rad C1000 Touch™ Thermal cycler with a Bio-Rad CFX96 Real-Time System (Bio-Rad, UK). Amplification consisted of 95°C for 2 min, followed by 50 cycles of 95°C for 10 s and 60°C for 1 min, with fluorogenic signals measured during annealing/extension step.

4.3.6 RT-qPCR data processing and statistical analysis

All data collection and analysis was performed in accordance with the The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin *et al*., 2009). Fluorogenic data were analysed using Bio-Rad CFX Manager Software v2.1 (Bio-Rad, UK). The quantification cycle value (Cq) was determined using the single threshold method, with the 'baseline subtracted curve fit' setting applied. Calibration curves were generated using the software's log-linear regression model after outlier exclusion. The coefficient of determination (R^2) and amplification efficiency (E = ($10^{(1/\text{-slope})}$ − 1) were calculated for each pathogen's individually-fitted calibration curve. Only calibration curves with R^2 0.9-1.0 and E 0.9-1.1 were kept for analysis. Sample Cq value was taken as an average of technical replicates, and the calibration curve for each assay was used to quantify copy numbers for each pathogen in the sample. DNA extraction losses and PCR inhibition for each sample were approximated by calculating Cq differences of amplified internal control DNA:

$$\%\, recovery = 1 - \left(\frac{Cq_{\max} - Cq}{-30}\right) x\ 100\% \qquad (2)$$

Whereby Cq_{max} is the highest internal control DNA Cq value observed, when averaged over both technical replicates (assumed to represent highest extraction and PCR efficiency) and Cq is the Cq value for the internal control DNA for each sample, averaged over both technical replicates. Copy numbers were corrected for extraction losses and PCR inhibition by dividing by the percent recovery calculated in Equation 2.

A generalised linear mixed model (GLMM) approach was used to evaluate the effect of different water treatments on pathogen copy numbers over time, using *R* v.3.4.3. GLMMs containing observation-level random effects (OLRE) were used to account for the overdispersion that is frequently observed in count data (Harrison, 2014), helping obtain a better fit of data to the model. This was achieved by adding a random effect to the model

that varied with each observation. GLMMs were run using the lme4 package (Bates *et al*., 2014) in *R* with the Poisson family and log link function used; the optimum model was selected using a downward stepwise procedure based on the AICc score (Burnham and Anderson, 2003). A two-stage (biexponential) die-off model as first described by Crane and Moore (1986) was used to derive die-off rate coefficients for each pathogen under the different water treatments. In similar studies examining die-off of bacterial indicator organisms, a biphasic decay of a fast, followed by a slower extended decay is frequently observed (Pachepsky *et al*., 2006; Rogers *et al*., 2011; Brouwer *et al*., 2017). The biexponential model consists of two first-order decay equations to encompass both stages of decay:

$$N_t = N_0 e^{-\mu_1 t} \quad when \quad t < t_a \qquad (3)$$

$$N_t = N_0 e^{-\mu_1 t_a} e^{-\mu_2 (t - t_a)} \quad when \quad t \geq t_a \qquad (4)$$

whereby N_t is the pathogen copy number at time *t,* N_0 is the initial inoculation copy number (averaged over the three experimental replicates), μ_1 and μ_2 are first-order decay rate constants (copy number day^{-1}) for the first (fast) and second (slower) decay stages respectively, and t_a is the time (days) when the first decay stages ends. The decay rate constants were derived by performing a series of non-linear regressions for each treatment (in *R*) making gradual iterations to the μ_1 coefficient and assessing the corresponding goodness of fit (Pearson's correlation coefficient). Optimised coefficients were used to fit the data to two regression models; one for time points ≤7 days and one for time points ≥14 days. The t_a value was calculated as the intersect of the two regression models. Generalised linear models (GLMs) were also used to evaluate the effect of water treatment and pathogen species on μ_1, μ_2 and t_a parameters. Probability values of <0.05 were considered statistically significant. Figures were generated by using the ggplot2 package (Wickham, 2016) in *R*. Observed vs modelled data were visualised in ggplot2 using *R*.

4.4 RESULTS

4.4.1 DNA extraction and qPCR amplification quality control

Internal amplification control DNA was successfully recovered from all spiked soil DNA samples. No target DNA was amplified in any NTC reaction (Table 4.2), indicating the absence of extraneous DNA molecules. Coefficient of determination (R^2) and amplification efficiency (E) values for all calibration curves were all within an acceptable range (Table 4.2), indicating effective DNA extraction and performance of the qPCR kits. Due to the low variation between the 3 technical replicates in the preliminary study, only 2 technical replicates were used for the longer-term survival experiments.

4.4.2 Indigenous pathogens and their regrowth

In non-spiked soils, indigenous *C. perfringens* was detected in 42% of samples (n = 12) at 60 - 2088 copies, *Cl. difficile* in 41.7% of samples (n = 12) at 24 - 2016 copies and *C. sporogenes* in 100% of samples (n = 18) at 20 – 3.27×10^4 copies g^{-1} of soil respectively. Pathogen copy numbers were not statistically different after 84 days, for any water treatment ($p > 0.05$).

4.4.3 Short-term C. sporogenes *persistence*

In the preliminary experiment looking at short-term survival of *C. sporogenes*, target copy numbers (from indigenous and spiked pathogens) declined over the 7-day sampling period in both high (flooded) and low (20% WHC) water treated soils (Figure 4.2). From an initial concentration of 6.4×10^5 copies g^{-1} soil ($n = 3$, *s.e.* = 6.2×10^4) at Day 0, after 7 days, concentrations had decreased to 4.28×10^5 ($n = 3$, *s.e.* = 3.12×10^4) and 5.56×10^5 ($n = 2$, *s.e.* = 7.6×10^4) copies per g^{-1} soil under high and low water treatments respectively. *C. sporogenes* copy numbers significantly

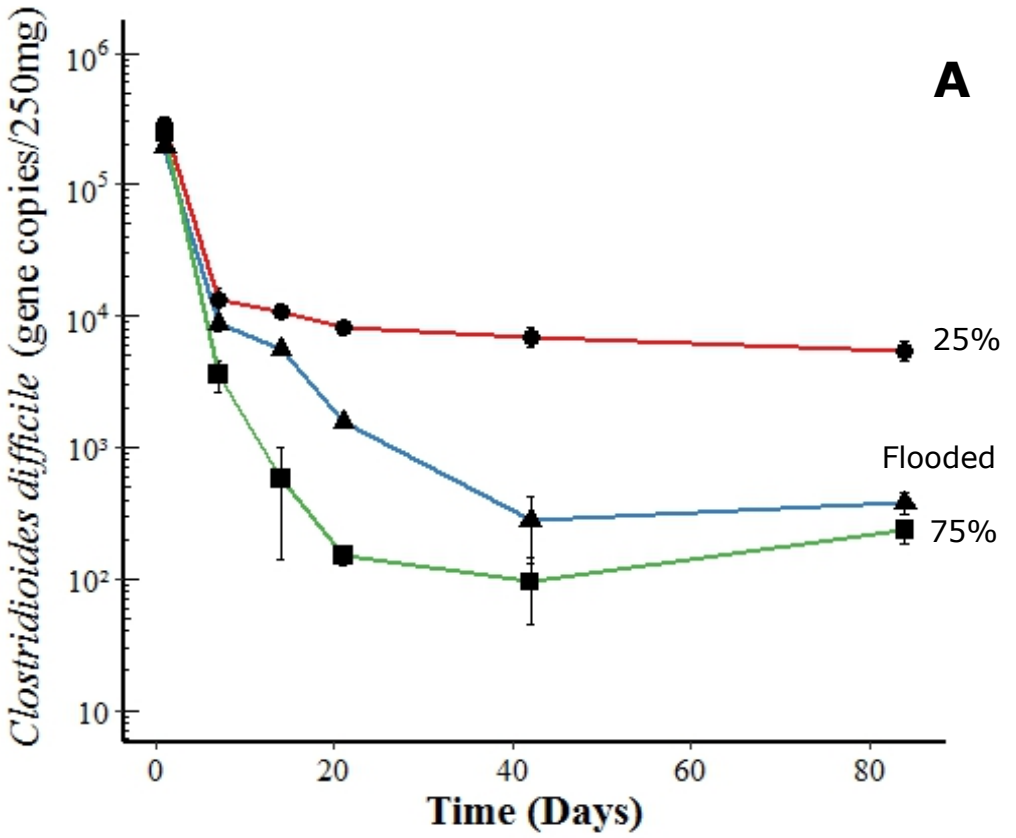
Table 4.2 Quality control parameters for RT-qPCR assays

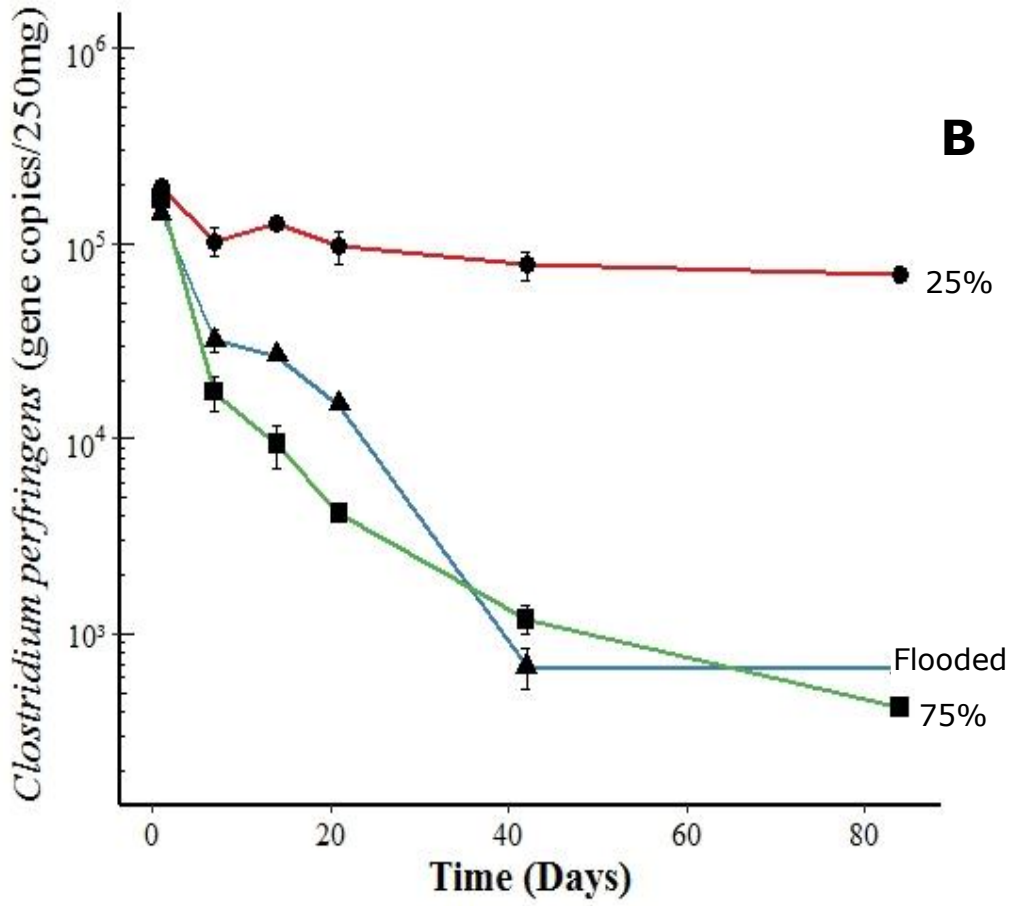
	Calibration Curve					IAC Quality Control	
	R^2	*E*	Outliers removed	Variation at LOD	NTC (*n=3)*	IAC variation	IAC fails
C. sporogenes (short term)	0.99	97.7%	1 (2 x 10^4 standard)	36.1 - 37.1	ND	28.7 - 32.4	0
Cl. difficile	0.97	94.5%	none	36.6 - 37.1	ND	NA	0
C. perfringens	0.99	96.2%	none	36.4 - 38.3	ND	28.7 - 30.5	0
C. sporogenes	0.97	101%	none	36.9 - 40.2	ND	NA	0

decreased over the seven days ($z = -4.17$, $p < 0.001$), and were higher in soil samples under the low water treatment ($z = 3.16$, $p = 0.0016$).

4.4.4 *Long-term persistence of* Cl. difficile*,* C. perfringens *and* C. sporogenes

Pathogen numbers decreased rapidly in a biphasic manner for all three species under every water treatment over the 84-day sampling period (Figure 4.2). Generally, increasing time after spiking led to a significant reduction in copy numbers for *Cl. difficile* ($z = -6.35$, $p < 0.001$*)*, *C. perfringens* ($z = -9.78$, $p < 0.001$*)* and *C. sporogenes (*$z = -4.59$, $p < 0.001$*)* respectively. Bacterial die-off was most rapid over the first seven days after spiking for all pathogens and soil water treatments, consistent with the two-stage die-off model. The initial decay (0-7 days) was greatest for pathogens under the 75% WHC water treatment, where 0.99 - 1.98 log copy number reductions were observed. Interestingly, pathogen decay was decreased in soils under the drier 25% WHC treatment. Copy numbers were significantly higher in the drier soil than those under wetter conditions for *Cl. difficile* ($z = 2.59$, $p = 0.01$*)*, *C. perfringens* ($z = 2.32$, $p = 0.021$*)* and *C. sporogenes (*$z = 2.05$, $p = 0.04$*).* There was no statistical difference between copy numbers in the 75% WHC and flooded treatments for *C. perfringens* ($z = -1.54$, $p = 0.123$*)* or *C. sporogenes (*$z = -1.60$, $p = 0.11$*),* although copy numbers of *Cl. difficile* were lower in the 75% WHC soil than in the flooded soils ($z = -2.13$, $p = 0.034$*).* The interaction between time and the driest water treatment was also significant for *C. perfringens* ($z = 5.88$, $p < 0.001$*)* and *C. sporogenes (*$z = 1.99$, $p = 0.047$*).*


A
Clostridioides difficile (gene copies/250mg)
10^6
10^5
10^4
10^3
10^2
10
0
20
40
60
80
Time (Days)
25%
Flooded
75%


B
Clostridium perfringens (gene copies/250mg)
10^6
10^5
10^4
10^3
0
20
40
60
80
Time (Days)
25%
Flooded
75%

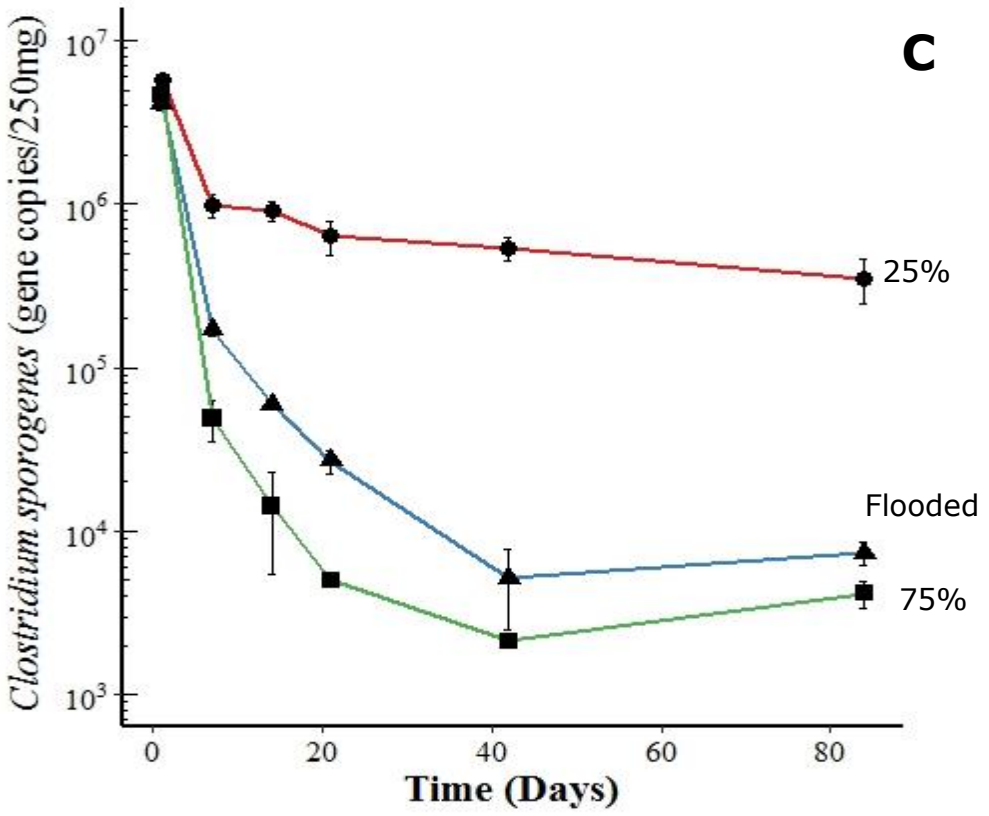


Figure 4.2 Measured decay of qPCR markers for *Cl. difficile* (**A**), *C. perfringens* (**B**) and *C. sporogenes* (**C**) after spiking into soil microcosms under 25% WHC (red line/circles), 75% WHC (green line/squares) or flooded (blue line/triangles) water treatments. Error bars show standard error for triplicate microcosms.

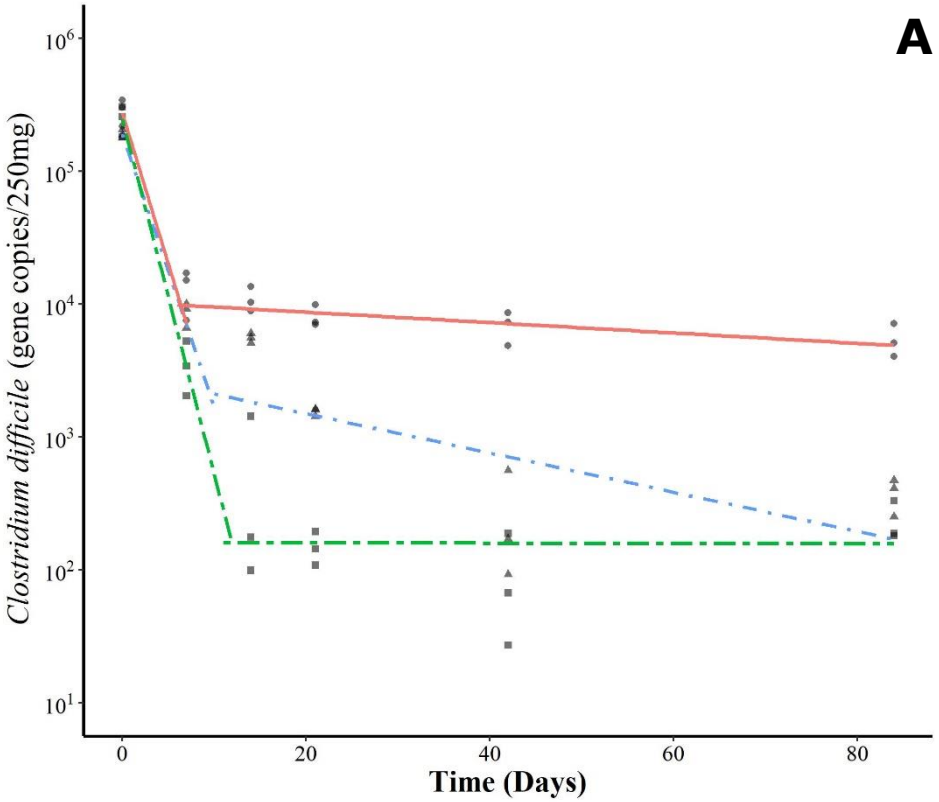
First-order decay rate constants, μ_1 and μ_2, and the transition time between the two phases of the decay, t_a, were derived for each pathogen and experimental condition (Table 4.3). Initial decay constants (μ_1) were considerably greater than decay rates for the second phase of decay (μ_2). *Cl. difficile* decreased at 0.47 - 0.62 d^{-1}, *C. perfringens* at 0.11 – 0.34 d^{-1}, and *C. sporogenes* at 0.35 – 0.67 d^{-1}. For all pathogens, t_a values in dry soils were approximately half that of the wetter soils. For all pathogens, second-stage decay constants were less than their first-stage values and were all less than 0.053 d^{-1} for all water treatments.

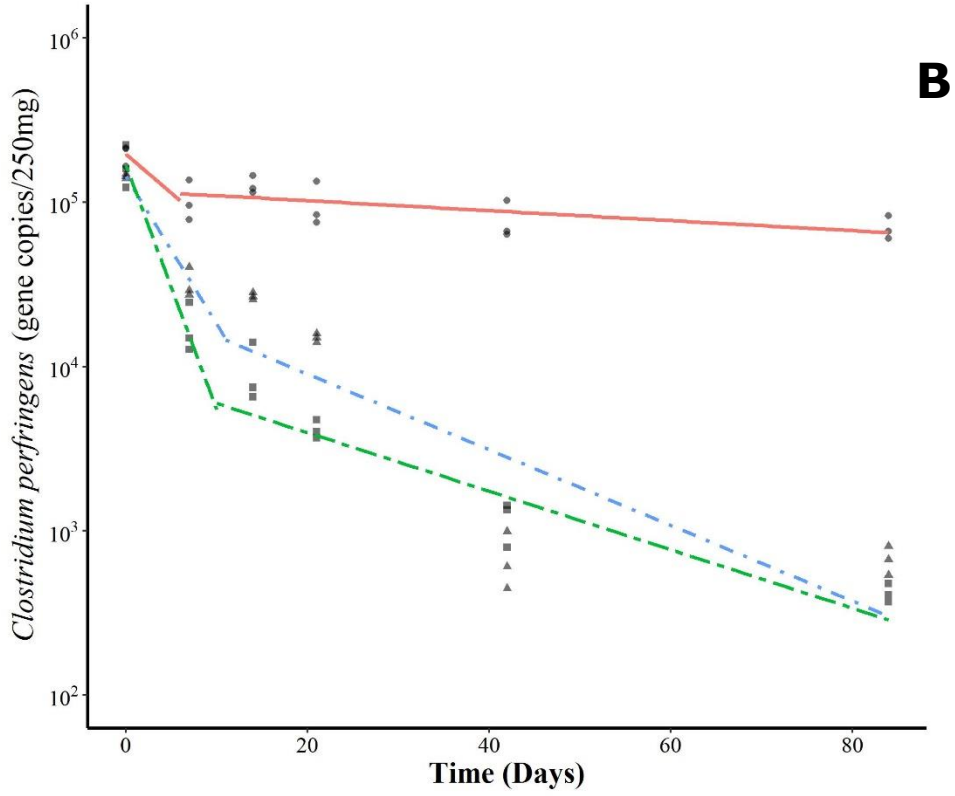
Table 4.3. Kinetic parameters of two-stage decay models applied to pathogen copy numbers

Pathogen	WHC[1]	μ_1	μ_2	t_a	R^2	p
Cl. difficile	Low (25%)	0.56	0.01	6.04	0.96	<.001
	75%	0.62	<.001	11.9	0.95	<.001
	High (flooded)	0.47	0.03	9.48	0.99	<.001
C. perfringens	Low (25%)	0.11	0.01	5.04	0.77	<.001
	75%	0.34	0.04	9.72	0.91	<.001
	High (flooded)	0.21	0.05	11.1	0.98	<.001
C. sporogenes	Low (25%)	0.35	0.01	5.48	0.97	<.001
	75%	0.67	0.01	10.3	0.92	<.001
	High (flooded)	0.47	0.02	11.2	0.98	<.001

[1] WHC refers to the water holding capacity at which the microcosms were held.

Both pathogen type and water treatment exerted observable effects on best-fit decay coefficients and transition times. Pathogen species affected the magnitude of first-stage decay rate coefficients ($\mu 1$), in the order *C. perfringens* < *C. sporogenes* (t = 4.67, p = 0.003*)* < *Cl. difficile* (t = 5.56, p = 0.005), although had no observable effect on μ_2 or t_a. Decreasing soil moisture to 25% WHC reduced the magnitude of $\mu 1$ (t = -3.43, p = 0.026) and t_a (t = -5.86, p = 0.004), although there were no significant changes in magnitude of μ_2 for any species/treatment combination. Figure 4.3 shows best-fit modelled biphasic decay curves over individual observations.


A
Clostridium difficile (gene copies/250mg)
10^6
10^5
10^4
10^3
10^2
10^1
0
20
40
60
80
Time (Days)


B
Clostridium perfringens (gene copies/250mg)
10^6
10^5
10^4
10^3
10^2
0
20
40
60
80
Time (Days)

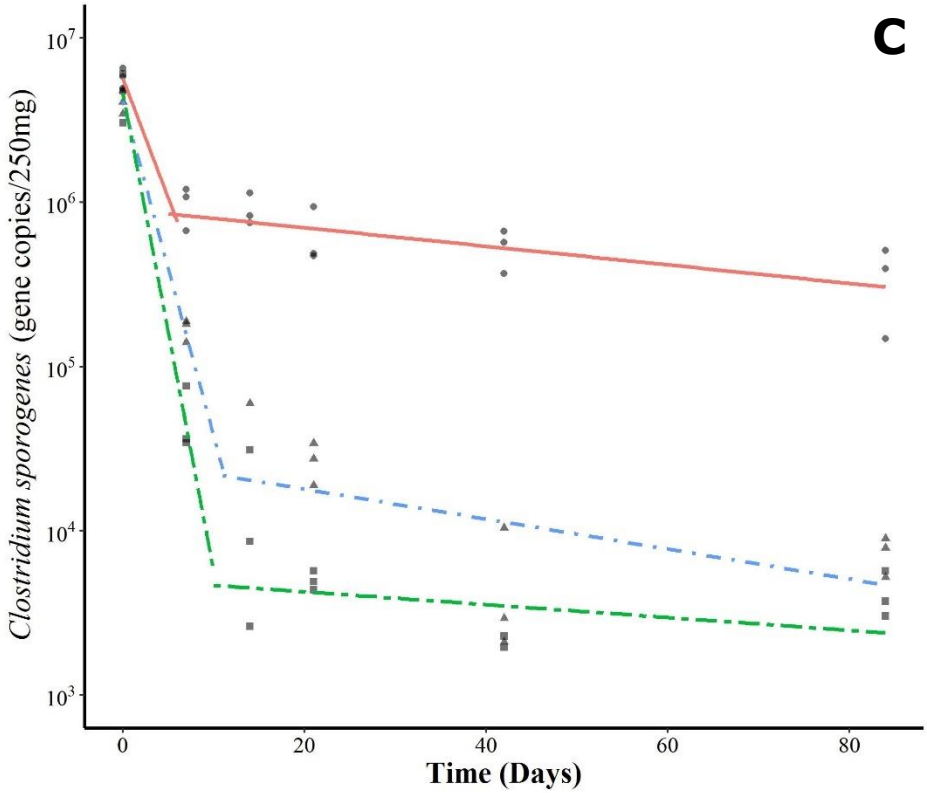


Figure 4.3 Modelled two-stage (biphasic) first-order decay curves for *Cl. difficile* (**A**), *C. perfringens* (**B**) and *C. sporogenes* (**C**) under 25% water holding capacity (WHC) (top, red dot-dash line), 75% WHC (bottom, green long-dash line) or flooded (middle, blue solid) water treatments. Models are overlaid on original data points.

4.5 DISCUSSION

The drivers of survival of clostridial pathogens in soils are not well researched, especially in comparison with other faecal indicator bacteria. As the clostridial pathogens studied are anaerobic, it was hypothesised that the soil moisture would likely be a key factor in moderating their survival, with wetter soils creating anaerobic conditions and thus favouring growth. This theory is supported by the high prevalence of Clostridia reported in anaerobic rice paddy soils (Weber, Stubner and Conrad, 2001; Liesack, Schnell and Revsbech, 2000; Liu *et al*., 2009) and other waterlogged soils (Rice and Paul, 1972; Pett-Ridge and Firestone, 2005). To this end, the objective of this study was to investigate the effect of soil moisture on the survival of the clostridial pathogens *Cl. difficile*, *C. perfringens* and C.

sporogenes (surrogate for proteolytic *C. botulinum*) in a typical agricultural soil.

4.5.1 Persistence of pathogens in soil

RT-qPCR was used to quantify genetic markers for the three pathogens over an eighty-four-day period, using soil microcosms. RT-qPCR was used in preference to culture-based techniques due to reported higher sensitivity and reproducibility, plus the ability to indiscriminately detect genetic markers originating from both vegetative and spore cells. The protective spore coat can make it hard to lyse spores comparatively to vegetative cells, which could lead to a reduction in gDNA yield from spores after extraction. To circumvent this problem, an additional freeze-thaw step was added (as recommended by the manufacturer) in the PowerSoil kit extraction protocol to improve spore lysis. This kit has also been shown to perform well in extracting gDNA from *Bacillus* spores, in comparison to other commercially available kits (Dineen *et al*., 2010; Sedlackova *et al*., 2017). Rogers *et al*. (2011) carried out a comprehensive study of culture-based versus RT-qPCR-based measurements of pathogens in manure-amended soils and found RT-qPCR markers to be reliable surrogates for culture-based techniques (Rogers *et al*., 2011). The RT-qPCR markers demonstrated predictable relationships to measurements from culture-based techniques over variable environmental stressors and time, warranting their use in this study. RT-qPCR is also capable of detecting live *C. perfringens* cells which could not be cultured using conventional methods (Matsuda *et al*., 2007).

Soil microcosms were amended with artificial rainwater to obtain soil resembling a realistic range of field WHC. To ensure soil structure homogeny and that all microcosms were amended to the correct water concentrations, it was necessary to air-dry soil overnight at 30°C prior to microcosm construction. This method of soil processing optimised accurate treatment assignment whilst minimising detrimental effects on the indigenous soil microbiology and chemistry. All pathogens persisted at

detectable levels for the duration of the study and were all present at concentrations of $>10^2$ copies per 250 mg of soil after eighty-four days. The soil water concentration had a significant effect on the survival of all three pathogens, with the driest soil (25% WHC) appearing to strongly enhance pathogen survival (Figure 4.2). This observation is likely explained by the spore-forming behaviour of most clostridial bacteria. When exposed to the unfavourable growth conditions of the dry soil, vegetative cells produce environmentally-resistance spores which then remain at consistent abundance for the duration of the sampling period. Pathogens spiked into the wetter soils likely remain in a vegetative state, more susceptible to deactivation (Goh, Riley and Chang, 2005). The rapid die-off seen over the first day is likely to be due to loss of cells to shock of change from culture to soil, or possibly even by between-species inhibition or competition (Smith, 1975b). After the first period of rapid decay, pathogen numbers stabilised, and die-off was minimal. Only the pathogen *Cl. difficile* was detected at significantly higher numbers in the flooded soils than in the 75% WHC soil ($p = 0.034$), suggesting that seasonal flooding events may not enhance the survival or proliferation of all pathogens as hypothesised. One practical concern is the overestimation of bacterial numbers by detection of extraneous DNA free in the soil or from dead cells. One solution is to co-extract DNA with propidium monoazide (PMA) which inhibits the PCR amplification of markers from extracellular DNA and dead cells. Bae and Wuetz (2009) found that while the PMA-qPCR counts of *Bacteroides* were lower than those using qPCR, decay curves followed remarkably similar patterns. The detection and recovery of naked DNA following various extraction processes has shown to be very low (Frostegård *et al*., 1999). Therefore, it can be assumed that pathogen DNA detected and quantified is reflective of the true number of live vegetative or spore cells present at the time of sampling. From a health perspective, as both cell and spore forms can induce illness, differentiation between the two is not necessitated. This study sought to identify the relationship between soil water and clostridial pathogen populations. However, future studies may complement RT-qPCR with culture-based approaches,

allowing a more comprehensive insight into the complex behaviour of spore-forming pathogens in soil.

4.5.2 Modelling pathogen decay

A biphasic decay model was utilised to understand the kinetics of the pathogen decay after spiking into the soil, producing useful die-off coefficients for predicting the environmental persistence of the pathogens investigated. Such models have been successfully used to describe the two clear stages of decay observed in the data and by most bacteria after introduction into an environment (Rogers *et al*., 2011; Brouwer *et al*., 2017; Bae and Wuertz, 2009; Crane and Moore, 1986). Using the biphasic model (Equation 4) die-off coefficients for the first, rapid period of decay (μ_1) and second, slower phase (μ_2) were derived for each pathogen and water treatment, and the transition time between the two phases (t_a) (Table 4.2). These predictors were then used to model decay curves for each pathogen under the different water treatments (Figure 4.3), which gave a good fit to the observed data. In the drier soils, not only was the initial decay rate significantly less, but pathogen populations transitioned into the second, slower phase of decay much earlier (5-6 days). This is indicative of the organisms entering spore-state upon exposure to unsuitable growth conditions. There was little discernible difference between pathogen die-off dynamics under the two wetter soil treatments, although pathogen species affected the μ_1 coefficient. The latter decay rate (μ_2) was unaffected by species or water treatment, indicating that pathogens reached a stable carrying capacity. These kinetic values agree with the few, comparable studies which have reported decay rate constants for Clostridia. *C. perfringens* first-stage decay constants (μ_1) in sand were reported as 0.59 d^{-1}, with t_a of 3.5 days and μ_2 0.04 d^{-1} (Zhang, He and Yan, 2015). Girardin *et al*., (2005) used a monophasic, first-order model to derive the decay of *C. sporogenes* in compost amended soil, which decayed at 0.007 log d^{-1} over 16 months, which is consistent with the μ_2 reported in this study. The decay coefficients derived in this study will

provide a basis for the development of quantitative microbial risk assessment models for predicting the residence of clostridial pathogens in agricultural soils after application of contaminated organic material. However, as an immediate consequence of this study, it seems that clostridial pathogen introduction to dry soils gives rise to significant long-term soil contamination. *Clostridium* spores contaminate overlying crops grown in manure amended soils (Östling and Lindgren, 1991) and this may compromise the safety of produce, feedstock and silage prepared from the crop. Transfer of dry soil, particularly *via* soil dust, may also present a significant health risk.

4.6 CONCLUSIONS

The results of this study demonstrate the ability of *Cl. difficile*, *C. perfringens* and C*. sporogenes* to persist in an agricultural soil at $>10^2$ gene copies 250 mg^{-1} soil. Decay was significantly reduced in dry soils (25% WHC), where populations $>10^4$ gene copies per 250 mg soil were still detected after eighty-four days. In dry soils, these clostridial pathogens rapidly transition to a stage of very slow decay after 5-6 days, after which decay rates are minimal. Slurry, manure or bio-solid application to dry soils may result in elevated concentrations of these pathogens which could pose a risk to human health. Future research is needed to investigate the complex behaviour in spore-forming pathogens in soil, particularly how different soil types and properties, and environmental conditions can impact pathogen decay, the timing of spore-formation and the subsequent risk to human and animal health. The decay coefficients derived in this study provide a basis for predicting long-term clostridial pathogen contamination, although more consideration of different soil factors influential to pathogen survival will allow refinement of coefficients for different soil types.

CHAPTER 5: DEVELOPMENT AND FIRST EVALUATION OF NOVEL AMRISA PROTOCOL FOR RAPID, SIMULTANEOUS DETECTION OF CLOSTRIDIAL PATHOGENS

5.1 ABSTRACT

The need for an automated diagnostic tool allowing identification and detection of multiple clostridial pathogens, including *Clostridioides difficile, Clostridium perfringens, C. botulinum* and *C. tetani* was identified. The limitations of existing molecular assays are discussed, including time, sensitivity and resolution issues; ergo, a suitable new and novel protocol was devised and evaluated. Newly available genomic data were utilised to identify suitable markers for PCR-based detection. Three protocols were designed *in silico* that utilised either terminal-restriction fragment length polymorphism (T-RFLP) or automated multiplex ribosomal intergenic spacer analysis (AMRISA) to detect and distinguish between clinically important species and strains. The T-RFLP protocols targeted the *pgk* gene, and while showing good specificity to target pathogens, could not be fully optimised due to poor sensitivity and reproducibility. The AMRISA protocol targeted the 16S-23S rDNA intergenic transcribed spacer (ITS) region. This test facilitated the direct identification (*in vitro*) of seven test clostridial pathogens from DNA of pure strains and mixed-species templates using easily-interpretable electropherograms. AMRISA was then used to identify contamination in test soils and anaerobic digestates, demonstrating the potential applicability of the assay for environmental pathogen surveillance or routine screening of complex clinical, food or environmental samples. Genomic DNA dilutions and live-cell-dilution lysates were both used to estimate the limits of detection for the protocol. Contamination as low as ~500 genome copies or lysate from 1.17×10^3 CFU per PCR reaction could be detected, although there is scope for improving the sensitivity. Furthermore, *in silico* analysis with the protocol on genomic sequences

revealed unique profiles for different clinical strains and pathotypes of certain pathogens, which could aid in understanding the epidemiology of some hypervirulent clostridial strains. Future work should develop the protocol by testing on a wider pool of known clinical and environmental isolates, generating a comprehensive reference database for comparison. This will increase the applicability of the assay.

Keywords; *Clostridium*, Clostridia, pathogens, multiplexing, PCR.

5.2 INTRODUCTION

Pathogenic Clostridia are the aetiological agents for various significant human and animal diseases. Many pathogenic clostridial species are ubiquitous in natural environments such as soils, sediments, water-systems and as common constituents of the mammalian gut-microflora. This engenders various complex pathogen fluxes and interactions within and between these reservoirs, subsequently giving rise to multiple disease exposure routes. Additionally, the longevity of the environmentally-resistant spores further increases the risk of disease transmission, with clostridial contamination an issue in clinical (Aldape, Bryant and Stevens, 2006; Peterson *et al*., 2011; Persson, Torpdahl and Olsen, 2008), veterinary (Songer, 1996; Vidal *et al*., 2013) and food production (Carlin, 2011; Fach *et al*., 2009) settings. Clostridial infections and outbreaks can be polymicrobial in nature and caused by low-level contamination (Brook, 2016). Moreover, fast diagnosis and treatment implementation is often hampered by the slow-growing and fastidious nature of pathogen cultivation using conventional techniques (Brook, 2016). Understanding the underlying epidemiology of harmful bacteria within all reservoirs is therefore essential knowledge if aiming to reduce exposure and improve disease prevention and control. Access to robust, rapid and sensitive diagnostic methods that can detect, differentiate and quantitate key pathogenic Clostridia underpins future research and disease-prevention.

The shortcomings of existing culture, biochemical, and immunological-based detection methods for Clostridia are discussed by other authors (Dhalluin *et al*., 2003; Le Bourhis *et al*., 2005; Janvilisri *et al*., 2010; Cao *et al*., 2011). In brief, common disadvantages include the low-sensitivity, cumbersome and costly nature of tests, the long duration before results can be obtained, and the ethics of using laboratory animals for pathogen detection (Lindström *et al*., 2001; De Medici *et al*., 2009; Cao *et al*., 2011). Current diagnostic methods for screening environmental samples favour molecular approaches. Whole genome sequencing (WGS) is considered the most informative tool for epidemiological strain typing (Köser *et al*., 2012; Salipante *et al*., 2015). Rapidly falling costs and turnaround times for WGS make this an increasingly useful approach for many studies, although for large scale environmental screening or rapid clinically identification, this is not currently the best diagnostic (Köser *et al*., 2012). Polymerase chain reaction (PCR) based techniques such as real-time, single or multiplex PCR, or microarray assay are the current gold standard methods of pathogen detection. In combination with various downstream genotyping or sequencing approaches, an automated, high-throughput detection protocol for multiple-clostridial pathogens is both conceptually feasible and epidemiologically necessary. Many studies have utilised the 16S ribosomal RNA (rRNA) gene for diagnostic, diversity and phylogenetic analysis (Hung *et al*., 2008; Cordoba *et al*., 2001; Dhalluin *et al*., 2003); although problems arise in the use of this gene when differentiation of clostridial species is required (Dhalluin *et al*., 2003; Janda and Abbott, 2007), as discussed in Chapter 2. The high conservation of this gene results in close nucleotide sequence similarity across the class. Consequently, differentiation between clinically different strains of some pathogenic species is not feasible using 16S sequencing or by techniques exploiting sequence polymorphisms within the gene. Furthermore, even distinction between species can be problematic (Sasaki *et al*., 2000b; Gurtler, Wilson and Mayall, 1991; Dhalluin *et al*., 2003; Mignard and Flandrois, 2006). Mignard and Flandrois (2006) used 16S DNA sequencing to identify 683 clinical isolates but could only identify 83.1% to species level. As discussed

in Chapter 2, many phylogenetically similar clostridial species, such as *C. botulinum* Group I / *C. sporogenes* and *C. botulinum* Group III / *C. novyi C. haemolyticum* (see Chapter 3- Figures 3.3B and 3.7, respectively), demonstrate markedly different modes of pathogenesis, or lack thereof, and hence future detection methods must reflect these important clinical distinctions. Conversely, the wide nucleotide heterogeneity at genome level between target pathogens creates difficulties for selecting alternate genetic markers (Le Bourhis *et al*., 2005). Janvilisri *et al.* (2010) successfully designed a diagnostic microarray assay to differentiate between more than four pathogenic *Clostridium* species from inoculated stool samples; however the detection limit was only 10^4 CFU ml^{-1} (Janvilisri *et al*., 2010). It is probable that in many environmental reservoirs, pathogen concentrations may fall below this detection limit, yet still pose an unacceptable risk to human and animal health. Therefore, a more sensitive detection method is still required.

The increasing number of publicly available clostridial genomes allows the assessment of other genetic loci as alternatives to the 16S rRNA gene for pathogen detection. Housekeeping genes could be suitable candidates for PCR-based molecular assays, as orthologous genes will be present in all target pathogens. Also, such genes generally show lower evolutionary rates and higher sequence identities than non-housekeeping genes (Jordan *et al*., 2002), resulting in highly-conserved sequence regions necessary for primer design. Conversely, non-essential genes (such as toxin-coding genes), are often absent from some targets (Lawley *et al*., 2009; Sebaihia *et al*., 2007; Dhalluin *et al*., 2003). However, it is anticipated that some housekeeping genes will also contain non-conserved regions, or areas with a high single-nucleotide polymorphism density that are required for molecular distinction between closely related clostridial pathogens. Indeed, Dhalluin *et al.*, (2003) used polymorphism analysis of the *tpi* housekeeping gene to differentiate between 12 clinically important *Clostridium* spp. In an earlier paper, Song *et al.* (2002) used the natural length-heterogeneity of the internal transcribed spacer (ITS) region, situated between the 16S and

23S rRNA genes, to distinguish between isolates of 17 common human intestinal Clostridia species with gel electrophoresis. Although no reference to the sensitivity or the applicability for use on co-contaminated, complex samples for these methods are given, it is thought that a similar approach can be developed for detection purposes. The aim of this chapter was to identify other suitable genetic markers for improved genotypic identification of clinically important clostridial species, and to develop a sensitive, high-throughput protocol for rapid clostridial pathogen detection. Assay results must be easily interpretable and facilitate distinction between different strains/pathotypes of target species. The virulence and disease mechanism of some clostridial pathogens is dependent upon the pathotype, and therefore a test which allows this level of discrimination will be of greater clinical and epidemiological use. Importantly, the protocol must be able to detect low-level contamination in complex environmental samples, such as soils and sediment, without use of an enrichment step prior to DNA extraction.

This study evaluated the suitability of 212 genetic loci for the detection and differentiation of clinically important clostridial pathogens using non-sequencing-based molecular techniques. Three promising protocols, targeting two different loci were devised *in silico* and laboratory tested for proof-of-concept. One technique developed utilised automated multiplex ribosomal intergenic spacer analysis (AMRISA) targeted at Clostridia. The assay could identify test pathogens as pure isolates and demonstrated a low limit of detection (LOD). The assay was further verified on DNA extracted from spiked soil samples and used experimentally to screen for clostridial contamination in anaerobic digestates and an agricultural soil.

5.3 Materials and Methods

5.3.1 Target pathogens

A comprehensive search of medical and veterinary literature was conducted to identify clostridial species which have been regularly

documented as causative agents of human or animal disease and should be included as target pathogens. The nomenclature for each pathogen was updated using their current taxonomic standing derived from the phylogenetic analysis conducted in Chapter 3, cross-checked with the List of Prokaryotic names with Standing in Nomenclature (Parte, 2018). Target pathogens are listed in Table 5.1.

5.3.2 Gene selection

A list of 212 genes was compiled for initial evaluation. These comprised the 'minimal bacterial gene set' of 206 essential genes proposed by Gil *et al.*, (2004), four Clostridia-specific genes identified by Lawley *et al.* (2009), and the entire 16S - 23S rRNA region (including the internal transcribed spacer [ITS] region). Gene nucleotide coding sequences (NCS) were evaluated on the National Centre for Biotechnology Information (NCBI) gene database (ftp://ncbi.nlm. nih.gov/gene). Genes with NCS <750 base pairs (bp) in length were omitted from selection; additionally, genes not present in all clostridial representatives (after searching for the conventional gene abbreviation or protein function) were also omitted.

Seventy-three clostridial complete genomic sequences were used as templates to retrieve gene NCS (ftp://ncbi.nlm.nih.gov/genbank /genomes/bacteria). This comprised both pathogenic and non-pathogenic Clostridia. NCS for gene homologues were retrieved from all genomic sequences using BLAST using the 'discontiguous megablast' algorithm, and the 'align sequences' function with existing, non-redundant gene NCS (obtained from the NCBI Gene data base) used as search queries. NCS for each gene were exported from BLAST and aligned using MUSCLE

Table 5.1 Target clostridial pathogens and their toxins

Species	Toxins	Organism affected[1] /main diseases
Clostridioides difficile	TcdA (enterotoxin) TcdB (cytotoxin), various promoters	H & A, pseudomembranous enterocolitis (Songer, 1996)
C. baratii	BoNT[2] F	H & A, botulism (Popoff and Bouvet, 2013)
C. botulinum Group I (proteolytic)	BoNT A, B, F	H & A, botulism (Popoff and Bouvet, 2013; Weigand *et al*., 2015)
C. botulinum Group II (non-proteolytic)	BoNT B, E, F	H & A, botulism (Popoff and Bouvet, 2013; Stringer *et al*., 2013)
C. botulinum Group III	BoNT C, D C/D	A, botulism (Skarin and Segerman, 2014; Popoff and Bouvet, 2013)
C. butyricum	BoNT E	H, botulism, necrotizing enterocolitis (Meng *et al*., 1999; Popoff and Bouvet, 2013)
C. chauvoei	α-haemolysin, necrotoxin; β-deoxyribonuclease; γ-hyaluronidase; and δ-haemolysin	A, blackleg (Popoff and Bouvet, 2013)
C. haemolyticum	(*C. novyi* Type D)- β, θ	A, bacillary haemoglobinuria (Skarin and Segerman, 2014)
C. neonatale	A, B, C + β -haemolysin	H, neonatal necrotizing enterocolitis (Alfa *et al*., 2002)
C. novyi	Type A- α, γ, δ, ε	H & A, gas gangrene, bighead, necrotic hepatitis,

	Type B- α, β, ζ, η	bacillary haemoglobinuria (Popoff and Bouvet, 2013; Skarin and Segerman, 2014)
C. perfringens	Type A- α Type B- α, β, ε Type C- α, β Type D- α, ε Type E- α, ι	H & A, gas gangrene, colitis, food poisoning, necrotic enteritis, enterotoxaemia, enteritis, dysentery (Songer, 1996)
C. septicum	α-lethal, necrotic, haemolytic; β-deoxyribonuclease; γ-hyaluronidase; δ-O^2-labile haemolysin, collagenase	H & A, gas gangrene, malignant oedema, braxy (Popoff and Bouvet, 2013)
C. sporogenes	BoNT B2 [3]	H & A, botulism (Weigand *et al*., 2015)
C. tetani	Tetanospasmin	H & A, tetanus (Bruggemann *et al*., 2003)
Paeniclostridium sordellii	β-haemorrhagic toxin, lethal toxin	H & A, spontaneous endometritis, massive oedema, gangrene (Popoff and Bouvet, 2013)
Paraclostridium bifermentans	Fibrinolysin, lecithinase, and O2-labile haemolysin	H & A, gelatinous oedema, empyema, cervical lymphadenitis, shock (Biswas *et al*., 2018)

[1] Humans affected (H), animals affected (A)

[2] Botulinum neurotoxin. Toxin likely acquired by horizontal gene transfer from *C. botulinum* Group I strains

program (Edgar, 2004) with 16 iterations. Genes that could not be aligned, showed evidence of gene paralogy, or which had poor levels of conservation were discarded from the analysis. Aligned sequences were exported as multiple alignment files for primer design.

5.3.3 Primer design

All potential primer sets were obtained for each suitable gene by either manual identification or using *GEMI* (v.1.3.1) (Sobhy and Colson, 2012). This was achieved by using sequence alignment files for target-only species (Table 1). The 'find all the oligos in the consensus sequence' approach was used to design oligotides that were 17-25 bp in length with a maximum of 5 degenerate bases. Potential primer pairs which could yield an amplicon between 200-1500 bp were evaluated. This size amplicon corresponds to size standards available for use for fragment analysis. When unable to design single primer pairs for all targets, multiplex primer sets were considered. Each primer was then evaluated sequentially using an iterative workflow. Primers with less degeneracy were selected preferentially. The first primer specificity screen consisted of assessing any potential binding sites on the other non-target clostridial sequences. This was performed manually by screening primer sequences against non-targets using CLC Sequence Viewer (v. 7.6.1). Amplification of some non-target Clostridia was accepted, permitting it was possible to distinguish between targets and non-targets after fragment length analysis (Section 5.3.4). Primer pairs/sets which were specific to the target pathogen species were further evaluated for their specificity using MFEprimer-2.0 online software (Qu *et al*., 2012) for degenerate primers against the 'bacterial genome' database, or Primer-BLAST (Ye *et al*., 2012) for non-degenerate primers. A degenerate primer contains positions which contain several possible bases. Therefore, a degenerate primer mix will contain mixtures of similar, but not identical oligonucleotides. Primers pairs/sets which did not amplify non-target DNA were then checked for suitable % GC, self-complementarity, potential hairpin formation and melting temperatures (T_m). Based on these parameters, unsuitable primers were excluded. Amplicon sequences were downloaded for *in silico* downstream analysis. Candidate primers pairs/sets were tested against target genomes using *in silico* PCR capability of the FastPCR software (Kalendar *et al*., 2017).

5.3.4 In silico amplicon processing

Pathogen differentiation was achieved using either the natural amplicon length-heterogeneity (LH) or by using restriction digestion. By labelling forward and/or reverse primers with different fluorescent dyes, and if necessary, performing a restriction enzyme digestion step, single, pairs or patterns of fragment peaks specific to a target organism could be generated, as has been achieved in other studies (Elliott *et al*., 2012; Mukhopadhyay and Mukhopadhyay, 2007; Dhalluin *et al*., 2003). When restriction digestion was used, pathogens were identified using terminal-restriction fragments (TRF) with terminal-fragment length polymorphism (T-RFLP). When amplicon LH could distinguish between targets, labelled amplicons could be submitted directly for fragment analysis (FA). Fragment length and fluorescent dye determination was determined in an automated manner with the FA capability of sequencing machines.

Cut sites for all commercially-available restriction enzymes were mapped onto amplicon sequences using the 'Restriction Enzyme Analysis' function on CLC Sequence Viewer. Restriction enzymes which targeted different cut site regions on different targets were selected for analysis. Single, double or triple digests were performed virtually on the amplicon sequences with the Restriction Digest software on the Sequence Manipulation Suite at http://www.bioinformatics.org/ (Stothard, 2000). Full *in silico* PCR (+/- restriction digestion) + FA was then simulated with the best primer set/restriction enzyme combinations on all available target gene NCS, to calculate the protocol specificity for individual targets.

5.3.5 Bacterial strains and culture conditions

A total of 8 clostridial strains were used to test the primer sets. These were *Cl. difficile* NCTC 13307 (School of Medicine, University of Aberdeen), *C. perfringens* NCTC 13170, *C. sporogenes* NCTC 532 (Sigma Aldrich, UK), *C. novyi* NCTC 538, *C. septicum* NCTC 547 (T), *Paeniclostridium sordellii* NCTC 13356 (T) (ThermoFisher Scientific, UK), *C. tetani* NCTC 279 (T) and

Paraclostridium bifermentans NCTC 506 (Public Health England, UK) with a *Bacillus* sp. (James Hutton Institute, UK) used as a negative control (T denotes type strains). All clostridial strains were inoculated into 20 ml of thioglycolate broth with resazurine (TYGB) (Oxoid, UK) in sterile universal tubes from frozen or lyophilised cultures. Inoculations were performed in an MK3 Anaerobic Work Station (Don Whitley Scientific Ltd., UK). Anaerobic conditions were maintained by incubating in jars with anaerobic-gas generating kits (Oxoid, UK) at 37°C for 24 h. Five hundred µl of each culture was subcultured in triplicate into 20 ml of TYGB and incubated anaerobically at 37°C for 48 h. Three tubes were also inoculated with sterile deionized water (SDIW) as negative controls. After incubation, tubes were inverted, and 2 ml aliquots removed to sterile Eppendorfs, before centrifuging at 5,000 × *g* for 10 min at 4°C. The pellet was then washed in 1 ml sterile phosphate buffered saline (pH 7.4) by vortexing and centrifuging as above. This was repeated before resuspending the pellet in 500 µl of 1 × TE buffer (0.01 M Tris-HCL, 0.001 M EDTA) and storing overnight at 4°C before DNA extraction.

5.3.6 DNA extraction

Cell solutions were centrifuged at 5,000 × *g* for 5 min at 4°C, treated with 200 µl of 6 % Chelex ® 100 matrix (Sigma Aldrich, UK), vortexed, heated to 50°C for 20 min, vortexed again, then heated to 99°C for 8 min. Samples were immediately chilled on ice, centrifuged at 10,000 × *g* for 5 min at 4°C before harvesting the genomic DNA (gDNA) by carefully pipetting off the supernatant. Duplicate extracts were combined to provide sufficient quantities of template. To check gDNA extraction had been successful, 4 µl of gDNA were combined with 4 µl of 25% xylene cyanol (v/v in SDIW) (Sigma Aldrich, UK) and loaded into a 1% agarose gel (1% agarose (Bioline, UK) w/v in 1 × TBE buffer (Tris base (Promega, UK), boric acid (Promega, UK), 1 M EDTA (Sigma Aldrich, UK)). Before pouring gels, ethidium bromide was added to stain DNA. The gels were electrophoresed in a 1 × TBE running buffer before visualising under UV light with the

Genegenius Gel documentation system (Syngene, UK), and images captured with GeneSnap software (Syngene, UK). Band fragment size was compared to DNA size standard Hyperladder 1kb (Bioline, UK). If extraction was successful, gDNA was quantified using the Qubit ® dsDNA Broad Range assay kit (Invitrogen, ThermoFisher Scientific, UK).

5.3.7 Optimisation of PCR or mPCR protocol

Singleplex or multiplex PCR (mPCR) was performed using 25 µl reaction mixtures containing 5 ng of gDNA, 0.5-1 µl of both forward and reverse primers (10mM) (Applied Biosystems, UK), 0.5 µl of bovine serum albumin (BSA) (Roche, UK), 1 µl of deoxynucleoside triphosphate (dNTP) mix (20 mM), 2.5 µl of 10 × NH_4 Buffer and 1.25 U of Bio*Taq* DNA polymerase (Bioline, UK). Concentrations of $MgCl_2$ (50mM) (Bioline, UK), forward and reverse primers (10mM) (Sigma Aldrich, UK) were optimised. All PCR reactions were performed using an Eppendorf Mastercycler® Nexus gradient thermal cycler (Eppendorf, UK). Unlabelled primers were first used to check for successful amplification and to optimise PCR conditions. When downstream FA was to be performed, primers were 5' labelled using any of the following G5 set dyes; 6-FAM (6-carboxyfluorescein), VIC®, NED™ or PET® (Applied Biosystems, UK). Mixed target templates were also made containing gDNA from multiple species at 2 ng of DNA of each species per reaction. The reaction mixture was denatured at 95°C for 5 min, before 29-34 cycles of 94°C for 30 sec, annealing at temperatures ramped from 45 to 60°C for 30 sec, and extension at 72°C for 45-60 sec, followed by a final incubation at 72°C for 10 min. Further tests were carried out to optimize the annealing times and the required number of amplification cycles (data not shown).

PCR products were visualised on 1.5% agarose gel stained with ethidium bromide to check for successful amplification and to optimise PCR conditions. Band fragment size was compared to DNA size standard Hyperladder 100bp (Bioline, UK), and band intensity was used to estimate PCR amplicon yield using GelAnalyzer software (Lazar, 2010).

5.3.8 PCR limit of detection

The limit of detection (LOD) of the optimised PCRs was evaluated using two methods. Firstly, quantified gDNA for each species was normalised with molecular grade water to the equivalent of approximately 3×10^5 genome copies per PCR reaction. This was calculated with the NEBiocalculator™ (https://nebiocalculator.neb.com) using the genome length of the strain in nucleotide base pairs. LOD was determined using serial 10-fold dilutions of this as PCR templates (Fach *et al*., 2009). Secondly, a colony-PCR based method was used, based upon the method of Ruengwilysup *et al.* (2009). Briefly, 50 µl of *Pn. sordellii* cell culture from frozen stock was plated onto CP Chromoselect agar (Sigma Aldrich, UK) and incubated anaerobically at 44°C for 24 hrs. Ten *Pn. sordellii* colonies were aseptically removed and suspended in 500 µl of molecular grade water, before carrying out ten-fold serial dilutions. Cells were heat-lysed by boiling each dilution at 100°C for 15 min on a hotplate, chilling on ice and then centrifuging at 14,000 × g for 1 min at room temperature. A 10 µl aliquot of supernatant was then used as a template for the PCR reactions. Fifty µl of each dilution were used to determine viable counts, using the plating method described above. PCR was performed simultaneously on the same templates using a verified primer pair that target clostridial 16S rDNA (Clos16SF-5' CCTAGTGTAGCGGTGAAATGCG 3' and Clos16R- 5' CAGCCTRCAATCCGAACTGRG 3') to check for the presence of amplifiable gDNA. For both methods, 4 µl of PCR product were visualised on a 1.5% agarose gel stained with ethidium bromide to calculate the gDNA/CFU LOD.

5.3.9 Restriction enzyme digestion and fragment analysis

After PCR optimisation, cleaned or uncleaned labelled amplicons were both used for FA or T-RFLP. The effect of an ethanol-precipitation clean-up step (Dunbar, Ticknor and Kuske, 2000) on the resulting data-quality was assessed on a small subset of amplicons. Restriction digestion was based on Elliott *et al.* (2012), where either 50 ng or 200 ng of PCR product was

digested depending on whether a single-or-multiple species template was used. The digestion protocol was performed as per the manufacturer's instructions (New England Biolabs, USA) in 25 µl reaction mixtures. Digests of 2 h at 37°C followed by 10 min at 95°C were performed on an Eppendorf Mastercycler® Nexus gradient thermal cycler. For FA, 2 µl of either digestion product or PCR product (diluted to varying degrees in SDIW) were added to 11.8 µl of Hi-Di Formamide (ThermoFisher Scientific, UK) and 0.2 µl of GeneScan™ LIZ®-labelled 600 or 1200 bp size standard (Applied Biosystems, UK). Prior to FA, samples were denatured at 95°C for 5 min and then immediately chilled on ice for 5 min. FA was performed on an ABI PRISM 3730xl Genetic Analyzer (James Hutton Institute, Dundee, UK).

5.3.10 Test for PCR inhibition with environmental samples

To test for potential PCR inhibition and to assess the efficacy of the optimised protocols with 'real-life', complex samples, gDNA extracted from pathogen-spiked soils, agricultural soils and anaerobic digestates were used as PCR templates. Briefly, air-dried and sieved (< 2mm) soil was spiked with cultures of *Cl. difficile* NCTC 13307, *C. perfringens* NCTC 13170 and *C. sporogenes* NCTC 532 at approximately 10^6 CFU g^{-1} and DNA extracted, as described in Chapter 3, Section 4. Additionally, DNA was also extracted from clostridia-positive agricultural soil (Farm 4, Site H1 as detailed in Chapter 5, Section 3.1) and 2 anaerobic digestates. DNA extractions were performed as described in Chapter 5, Section 3.4. For each PCR reaction, 50 ng of gDNA was used.

5.3.11 Data analysis

Fragment data were imported into Peak Studio v. 2.2 software (McCafferty *et al*., 2012) for peak analysis and visualisation. This non-proprietary software was selected for its simplicity of use, easy inspection and manual manipulation of size standard peaks, and the ability to overlay sample profiles in one electropherogram. The software's built-in linear

interpolation peak-calling algorithm was utilised, and data smoothing was applied using the default options, as recommended by the author.

5.4 RESULTS

5.4.1 Gene selection and primer design

The sequence alignments for 47 suitable genes were imported into *GEMI* software for primer design. For these genes, conserved NCS were successfully retrieved and aligned for all target species; in addition to meeting the criteria detailed in Section 5.3.2. Candidate primers were designed for 35 genes given the specified settings, although from this pool, only 6 genes produced PCR-suitable primer pairs/sets. The main reason for gene elimination was the lack of dual conserved regions in genes required for >200 bp amplicon generation. Most primers designed were degenerate and complimentary forward and reverse primer pairs/sets with similar T_m were seldom found. Additionally, very low % GC was observed across virtually all genes analysed, which was prohibitive to the design of suitable primers. The 6 suitable genetic loci selected for further analysis were the 16S-ITS-23S rRNA region and *dnaK* (chaperone protein), *fusA* (elongation factor G), *pgk* (phosphoglycerate kinase) and *ychF* (ribosome-binding ATPase) genes.

5.4.2 In silico PCR and specificity checking

In silico PCR was simulated for all candidate primer pairs/sets for the 6 selected genes. *FusA* primers failed the first specificity test as they also amplified numerous non-clostridial sequences. While showing excellent specificity to targets, *dnaK* and *ychF* genes had insufficient genetic variation within amplicons to allow appropriate discrimination between target organisms, even when triple restriction digestions were simulated. These genes were eliminated from the analysis. Two different suitable genetic locations did allow for target pathogen discrimination. The *pgk*

gene showed highly conserved consensus regions flanking variable regions, for which species discrimination could be achieved after restriction digestion. Primer Pair A, consisting of *CpathPGKf* and *CpathPGKr* primers (Table 5.2), produced an 800 bp amplicon, containing variable regions satisfactory for enzymatic digestion for species differentiation. However, due to the low GC% of this gene, the primers had sub-optimal melting temperatures (T_m). Multiplex Primer Set B, also targeting the *pgk* gene, was also designed *in silico* (Table 5.2). This consisted of two degenerate forward primers: *ClospathAf* and *ClospathBf,* and two degenerate reverse primers: *ClospathXr* and *ClospathYr.* The T_m for these primers were all within the optimal range.

A third candidate primer set utilised the natural LH of the ITS region to distinguish between target bacteria. Two forward (*CpathAf* and *CpathBf)* and four reverse (*Cpath1r, Cpath2r, Cpath3r* and *Cpath4r),* non-degenerate primers were designed for multiplexing (Table 5.2). The forward primer *Cpath*Af targets DNA in the V8 hypervariable region of the 16S rRNA gene and *Cpath*Bf is a sequence just before the V9 region. The reverse primers target four pathogen group-specific regions on beginning of the 23S rRNA gene, and the entire ITS region of target pathogens is amplified. This technique, termed ribosomal intergenic spacer analysis, has been successfully used for bacterial community studies (Bulgari *et al*., 2009; Moreno *et al*., 2011), differentiation of species within a genus (Song *et al*., 2000) and for specific pathogen detection (Khan *et al*., 2007). The ITS region has also been used for strain and species-differentiation within clostridial species (Stubbs *et al*., 1999). *In silico* simulation of the full automated multiplex ribosomal intergenic spacer analysis (AMRISA) revealed that unique peak profiles were generated for different targets, with some profiles strain and pathotype-specific. This high-resolution differentiation power was not possible with Primer Sets A and B. Furthermore, the use of non-degenerate primers allowed for better primer T_m matching, reducing the likelihood of non-target amplification.

Table 5.2 Names and sequences of primers designed in this study

Set	Target and name	Sequence (5′ – 3′)	T_m (°C) [1]
A	***pgk* gene**		
	CpathPGKf	AGATAYAGAAAAGAAGARACTA	51.1 - 54.7
	CpathPGKr	CTTTWCCTTCTAARAATTCWAR	51.1 - 54.7
B	***pgk* gene**		
	CpathPGKAf	AGAGCWCAYTGCTCWACAGT	56.4 - 58.4
	CpathPGKBf	GATGAAAAYAGATTARWVGGAGC	55.5 - 60.9
	CpathPGKXr	GGTCCNATATCWAGTCCCAT	56.4 - 58.4
	CpathPGKYr	ATATCCTTGAGMTTTTARGAATG	53.9 - 57.6
C	**16S rRNA gene**		
	*Cpath*Af	TAACCGTAAGGAGCCAGCG	59.5
	*Cpath*Bf	TCTCAGTTCGGATTGTAGGCTGA	62.9
	23S rRNA gene		
	Cpath1r	CGGATCAAAGTTTACGTGCAACTC	63.6
	Cpath2r	ATTCACCATGCGCCCTTTGTAG	62.1
	Cpath3r	GCGCCCTTTGTAGCTTGATCTATC	65.2
	Cpath4r	CCGTTAGGTATGCAGGTTTCCC	64.2

[1] Melting temperature (salt adjusted) calculated using Oligo Calculator version 3.27 (Kibbe, 2007)

5.4.3 In silico restriction analysis

Restriction enzyme digestion was used to differentiate between amplicons of different target pathogens generated with Primer Sets A and B. For Primer Set A, a double digestion using *Sfc*I and *Ssp*I restriction enzymes produced the best discrimination between TRFs. For amplicons derived from Primer Set B, double digestion with *HpyCH*IV and *HypCH*V restriction enzymes gave the best discrimination. Corresponding TRFs for specific pathogens are shown in Table 5.3. For Primer Sets A and B, it was not possible to distinguish between Group I *C. botulinum* and *C. sporogenes*. As discussed in Chapter 3, phylogenetically, *C. sporogenes* is a non-toxic variant of Group I *C. botulinum,* and separate species status is not warranted. Similarly, existing nomenclature does not correlate with the correct phylogenetic lineages of *C. haemolyticum*, *C. noyvi* Type A and B and *C. botulinum* Group III strains.

5.4.4 Fragment database generation

For the AMRISA protocol using Primer Set C, a database was created for target strains using all publicly available genomic data. All positive hits from complete (fully assembled) genomes retrieved using primer-Blast are given in Supplementary Table 3. Numerous partially-assembled genomes and scaffold sequences for suspected target species were also screened directly using FastPCR software. Most pathogenic species could be identified by a unique peak labelled with one of four fluorescent dyes (Table 5.4), while for many pathogens it was also possible to differentiate between different strains and pathotypes. Other non-pathogenic species could also be detected and distinguished between.

Table 5.3 Terminal restriction fragments (TRFs) predicted for target species

Target	Primer Set A		Primer Set B		
	F-TF	**R-TF**	**F-TF[1]**	**R-TF[1]**	**# strains tested**
Cl. difficile	68	**215**	**A-25**	**X-73**	100
C. baratii	**711**	89	B-**239/343**	**X-316**	3
C. botulinum Gr. I	227	41	B-175	Y-152	42
C. botulinum Gr. II	**462**	292	**A-94** & **B-58**	Y-152	23
C. botulinum Gr. III	**495**/444[2]	155/359	**A-124**	X-70	24
C. butyricum	**482**	314	**B-58**	**Y-263**	8
C. chauveoi	68	292	ND	ND	2
C. haemolyticum	**491**	155	ND	ND	5
C. neonatale	**458**	292	**A-49**	**Y-197**	1
C. novyi	**440**	359	**A-341**	X-102	3
C. perfringens	**133/198**	314	A-161 & **B-188**	Y-85	16
C. septicum	66	293	A-161	Y-85	1
C. sporogenes	227	41	B-175	Y-152	4
C. tetani	117/**591**	41/89	A-40 & **B-322**	**X-64**	7
Pn. sordellii	68	89/116	B-175	X-40/70	38
Pr. bifermentans	68	41	B-175	X-40	3

[1] Forward TRFs (F-TF) and reverse TRFs (R-TF) length in base pairs. Unique TRFs in bold.
[2] Two alternate TRFs predicted *in silico.* For primer set B, specific forward and reverse primers are denoted by A or B and X or Y lettering, respectively. TRFs are listed using the code: labelled primer/ length (bp).

Table 5.4 Summary of peak-profiles for AMRISA targets derived from *in silico* analysis

Target species	Dye	Main peak(s)	Other peaks	Strains tested
C. acetobutylicum	VIC	456-7	466	3
C. baratii	VIC	530-3	560, 616,698	2
C. beijerinckii	VIC	499, 625	578, 615-7, 687	6
C. botulinum Group II	VIC	504-6	476, 656, 665, 862	6
C. butyricum	VIC	519-21/685	642	5
C. carboxidivorans	VIC	499-500	601, 672	1
C. chauvoei	VIC	507	584, 662	3
C. drakei strain	VIC	499	573, 663, 677, 846	1
C. perfringens	VIC	464	696-8	15
C. saccharobutylicum	VIC	520	555, 605, 621, 633, 683, 724	5
C. scatologenes	VIC	499	773	1
C. septicum	VIC	583	655, 734	1*
C. sp. MF28	VIC	499, 624	578, 615	1
C. taeniosporum	VIC	502-3	664, 745, 784-7	1
C. tyrobutyricum	VIC	517		1
C. botulinum Group I	VIC	344-7	281, 283, 345, 360, 498	40
C. carboxidivorans	VIC	349-50	451, 522	1
C. drakei	VIC	349	423, 696	1
C. scatologenes	VIC	349	623	1
C. sporogenes	VIC	350	422, 500, 559, 571	9
C. botulinum Group III	NED	275-8, 355, 424-8	446, 499	13
C. novyi Type A	NED	292	344, 429, 484, 529	2
C. novyi Type B	NED	279	429, 491, 641	2
C. cochlearium	PET	507	589, 773	1
C. tetani	PET	504, 588	787	3
Cl. difficile Clade I	6FAM	617, 639	577, 666, 737, 798, 859	7
Cl. difficile Clade II	6FAM	575, 637	542, 678, 735, 859	8
Cl. difficile Clade III	6FAM	595, 637	575, 756, 797, 859	3
Cl. difficile Clade IV	6FAM	679	542, 575, 619, 637, 735, 757	11
Cl. difficile Clade V	6FAM	618	639, 681, 723	4
Pn. sordellii	6FAM	555	496, 585, 603, 656-8	9
Pr. bifermentans	6FAM	501	518, 555	1*
Romboutsia ilealis strain CRIB	6FAM	570	499, 634, 641, 696	1
Romboutsia sp. Frifi	6FAM	558	495, 621, 696	1

* Peak profiles for these species were derived from testing on lab strains

5.4.5 Validation of (m)PCR-T-RFLP protocol

Genomic DNA was successfully extracted for all strains; except for *C. tetani* which could not be cultured. PCRs and mPCRs were optimised with gDNA templates of seven target species using amplicon visualisation with gel electrophoresis. An amplicon of ~800 bp was produced for Primer Set A, as predicted *in silico* (Figure 5.1). The PCR protocol for Primer Set B could not be optimised because the amplicon sizes predicted *in silico* were not reproducible with real strains. Multiple spurious bands were frequently observed (data not shown) illustrating the poor sensitivity of primer set B. For these reasons, Primer Set B was excluded from further analysis. The PCR protocol for Set A was further optimised using labelled primers. Optimal PCR conditions consisted of 25 µl reaction mixtures containing 2.5 µl of 10 × NH_4 buffer, 1.5 µl of 50mM $MgCl_2$, 0.5 µl of 20 mM dNTP mix 0.625 µl of each 5'-end 20mM 6-FAM labelled *CpathPGKf* and HEX™ *CpathPGKr* labelled primers, 0.5 µl of BSA and 1.25 U of *BioTaq* DNA. Cycling conditions were 95°C for 3 min, followed by 32 cycles of 94°C for 30 s, 47°C for 30 s and 72°C for 45 s, with a final extension of 72°C for 10 mins and storage at 10°C. For successful PCRs, 15 µl aliquots of PCR product were cleaned using ethanol precipitation, and the yield of DNA recalculated using Qubit. Fifty ng of cleaned or uncleaned amplicons were enzymatically digested with *Ssp*I and *Sfc*I restriction enzymes (New England Biolabs, USA), using 10 x Cutsmart buffer (New England Biolabs, USA) with the protocol recommended by manufacturer. Two µl of digestion product were submitted for fragment analysis (FA). When visualised, the level of agreement between predicted vs observed TRF was inconsistent, with differences of 1-8 bp observed. For some species, additional unexpected peaks were observed, and in some instances, predicted peaks were absent (data not shown). Digestion conditions were modified, and amplicons were resubmitted, but this did not resolve the issues encountered.

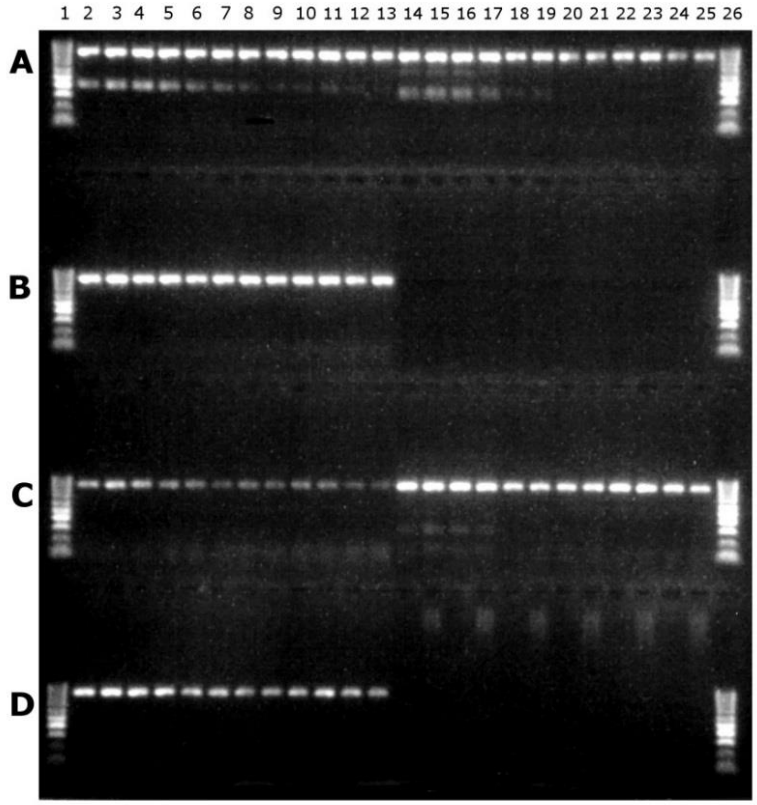


Figure 5.1 An example of T_m optimisation across a thermal gradient using primer set A. Gel shows an ~800 bp amplicon using gDNA templates of *Pr. bifermentans* (A2-13), *Cl. difficile* (A14-25), *C. perfringens* (B2-13), *C. novyi* (B14-25), *C. septicum* (C2-13), *C. sporogenes* (C14-25), *Pn. sordellii* (D2-13) and *Bacillus* sp. (D14-25). Lane 1 + 26: molecular weight marker (Hyperladder™ 100bp)

5.4.6 Validation of AMRISA protocol

Primer Set C demonstrated excellent sensitivity and specificity, with gel-visualised amplicons matching those predicted in silico (Figure 5.2). Final PCR conditions were 25 µl reaction mixtures containing 2.5 µl of 10 × NH_4 buffer, 2 µl of 50mM $MgCl_2$, 1 µl of 20 mM dNTP mix, 0.75 µl of CpathAf, CpathBf and 5'-end 6-FAM labelled Cpath1r, VIC™ labelled Cpath2r, NED™ labelled Cpath3r and PET™ labelled Cpath4r primers, each at 20 µM (Applied Biosystems, UK), 0.5 µl of BSA and 1.25 U of BioTaq DNA polymerase. Cycling conditions were 95°C for 3 min, followed by 32 cycles of 94°C for 30 s, 56°C for 30 s and 72 °C for 52 s,

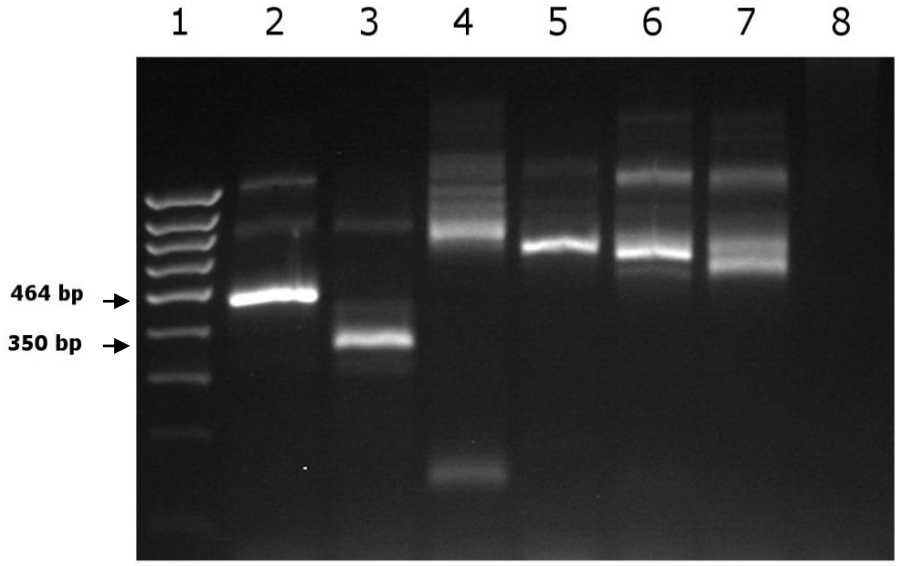
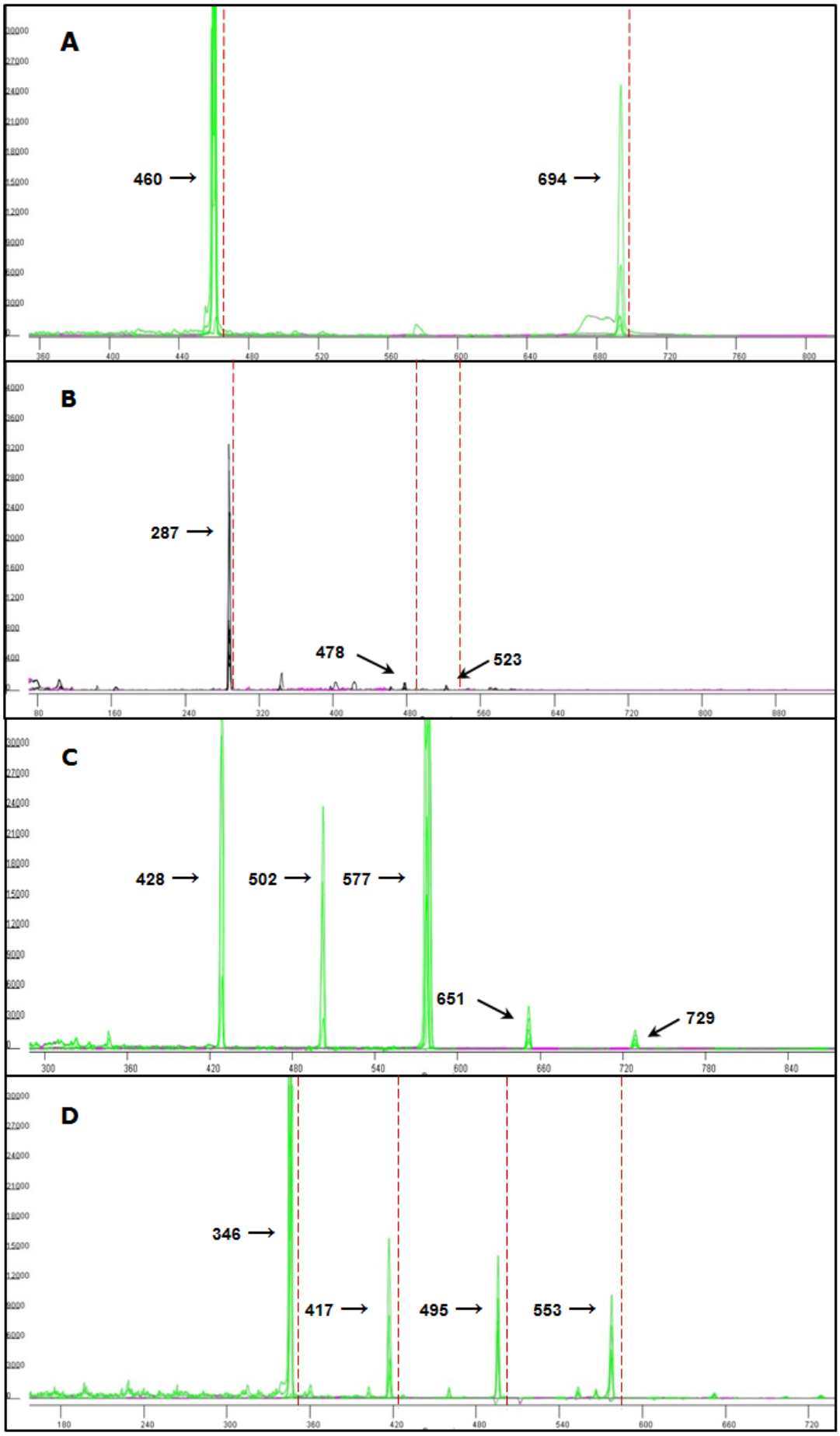


Figure 5.2 Multiplex PCR amplification products of primer set C. Amplicons from gDNA templates of *C. perfringens* (2), *C. sporogenes* (3), *Cl. difficile* (4), *C. septicum* (5), *Pn. sordellii* (6), *Pr. bifermentans* (7), and *Bacillus* sp. (8). Lane 1: molecular weight marker (Hyperladder™ 100bp).

with a final extension of 72°C for 10 mins and storage at 10°C. Amplicons for a subset of samples were cleaned using ethanol precipitation. For FA, 2 µl of cleaned or uncleaned amplicons (diluted 1:75 in SDIW) were submitted.

When visualised, fragment profiles were in good agreement with those predicted *in silico.* Consistent bp shifts were observed between predicted vs observed peaks; -8 to -10 bp for 6FAM peaks and -3 to -4 bp for other dyes (Figures 5.3 A-F). The electropherogram for fragments derived from a mixed-species template is shown in Figure 5.4. Species-specific peaks could be clearly identified for all species from the mixed-template sample. There was little discernible difference between electropherograms for cleaned and uncleaned PCR products (including for environmental samples), therefore this step was removed from the protocol.


A
460
694
B
287
478
523
C
428
502
577
651
729
D
346
417
495
553

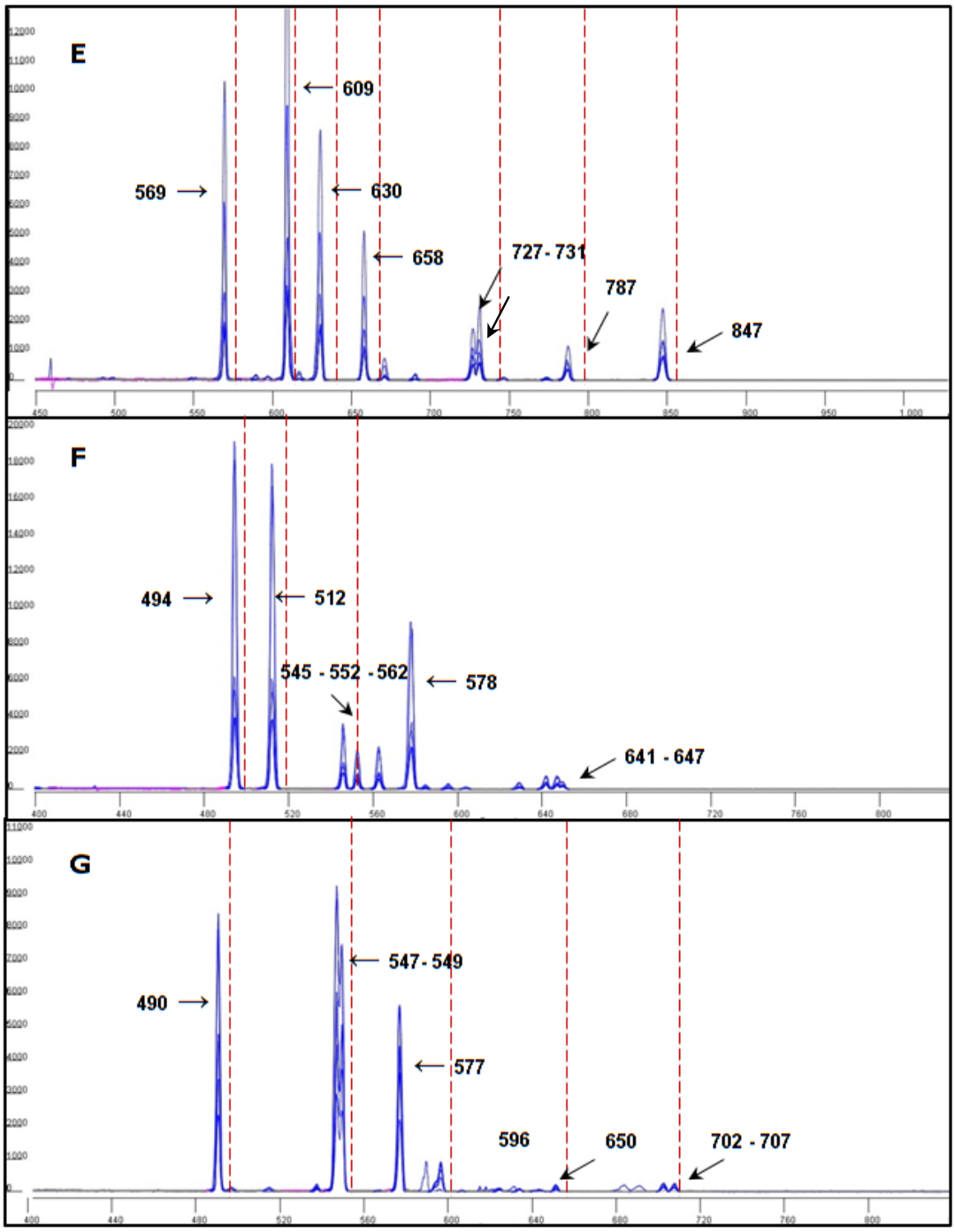


Figure 5.3 AMRISA profiles for pure isolates of *C. perfringens* (A), *C. novyi* (B), *C. septicum* (C), *C. sporogenes* (D), *Cl. difficile* (E), *Pr. bifermentans* (F) and *Pn. sordellii* (G). Observed peak sizes in base pairs are given, with peak colours indicating fragment dyes: 6FAM (blue), VIC (green) and NED (black). Y-axis shows relative peak fluorescence. Each profile represents 6 replicates for each species. Red-dashed vertical lines show the predicted size of peaks. Due to the lack of genomic data, peak sizes could not be predicted for *C. septicum.*

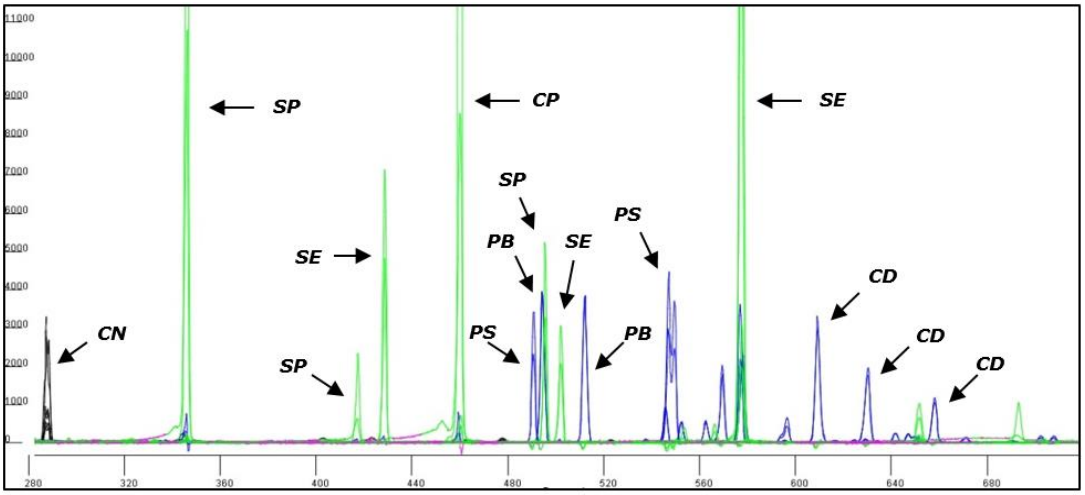


Figure 5.4 AMRISA profile for pre-amplification DNA template mixes. Template mixes containing 2ng of *Cl. difficile (CD), C. perfringens (CP), C. novyi (CN), C. septicum (SE), C. sporogenes (SP), Pn. sordellii (PS) and Pr. bifermentans (PB)* gDNA were used for mPCR reactions. 2 µl of amplification were diluted 1:75 in SDIW and submitted for fragment analysis. Profile shows results of four replicates.

5.4.7 Limits of detection

The LOD of the AMRISA protocol using Primer Set C was evaluated. To determine the LOD of the mPCR step, serial dilutions of gDNA were created for each species (except *C. novyi* due to very low extraction yield), corresponding to a range between 5×10^4 to 5 genome copies per reaction. PCR amplicons for the lowest four dilutions were visualised using gel electrophoresis (Figure 5.5). All mPCR reactions for *Cl. difficile* and *C. septicum* failed, for undetermined reasons. The LOD for all other species was approximately 500 genomes per reaction. Secondly, lysates prepared from serial dilutions of *Pn. sordellii* cells were used as a template for the mPCR, plus a PCR using pre-validated 16S Clostridia primers to check for the presence of amplifiable gDNA. Amplicons were detected using gel electrophoresis. The LOD for the mPCR was approximately 1.17×10^3 CFU per reaction, although the clostridial 16S rDNA primers yielded faint amplification products at 11.7 CFU per reaction (Figure 5.6), indicating amplifiable DNA was still present.

5.4.8 Application of AMRISA protocol on environmental samples

AMRISA was used to screen DNA extracts from pathogen-spiked soils and Clostridia-contaminated soils and anaerobic digestates. When MPCR amplicons were run on agarose gels, bands were visualised for all the environmental samples screened, indicating clostridial contamination sequences had been amplified. Amplicons were diluted in SDIW and submitted for FA. The resulting AMRISA profiles for selected dye traces are visualised in Figure 5.7, with matching pathogen traces overlaid. A tabular summary of the main peaks and putative pathogen matches is given in Table 5.5. Peaks for the pathogens were clearly identified in the spiked soil (Figure 5.7A, Table 5.5), while suspected pathogen contamination could also be identified in farm soil (Figure 5.7B) and both anaerobic digestates (Figures 5.7C + D).

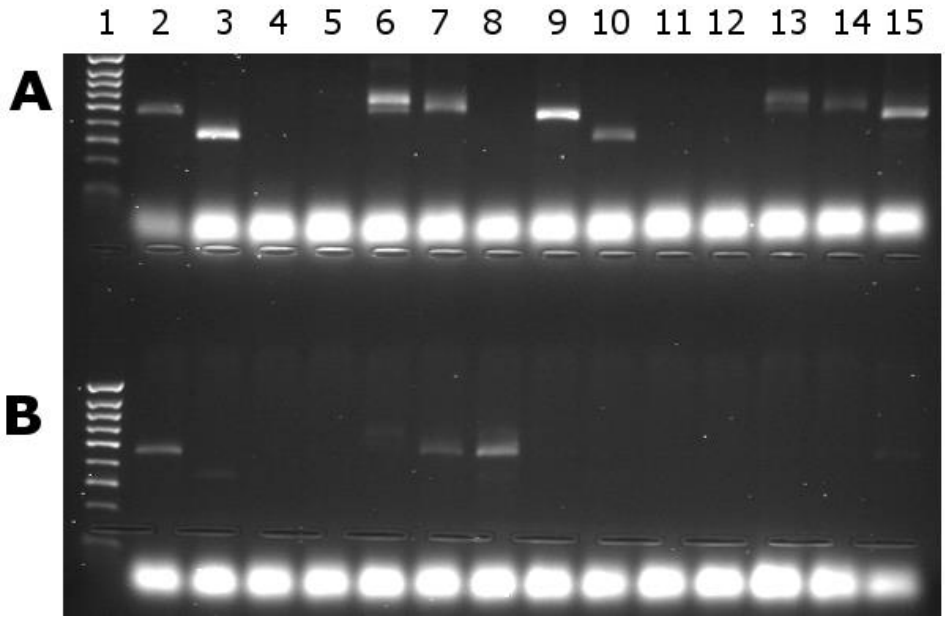


Figure 5.5 Multiplex PCR amplification products of AMRISA protocol with 10-fold serial dilutions of gDNA templates. The sensitivity of the mPCR step of the AMRISA protocol was assessed using 10-fold serial dilutions of gDNA equivalent to approximately 5 x 10^4 (A2-A8), 5000 (A9-A15), 500 (B2-B8) and 50 (B9-15) genome copies. Lanes: 1, molecular weight marker (Hyperladder™ 100bp); 2 + 9 *C. perfringens;* 3 + 10, *C. sporogenes*; 4 + 11, *Cl. difficile;* 5 + 11, *C. septicum*; 6 + 13, *Pn. sordellii*; 7 + 14, *Pr. bifermentans* and 8 + 15, mixed-template (8 & 15). Reactions for *Cl. difficile* and *C. septicum* all failed.

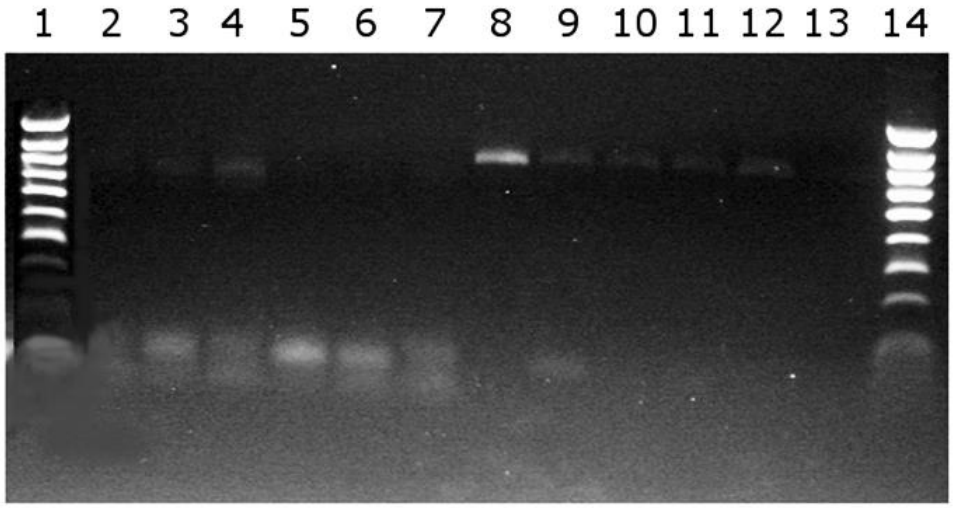
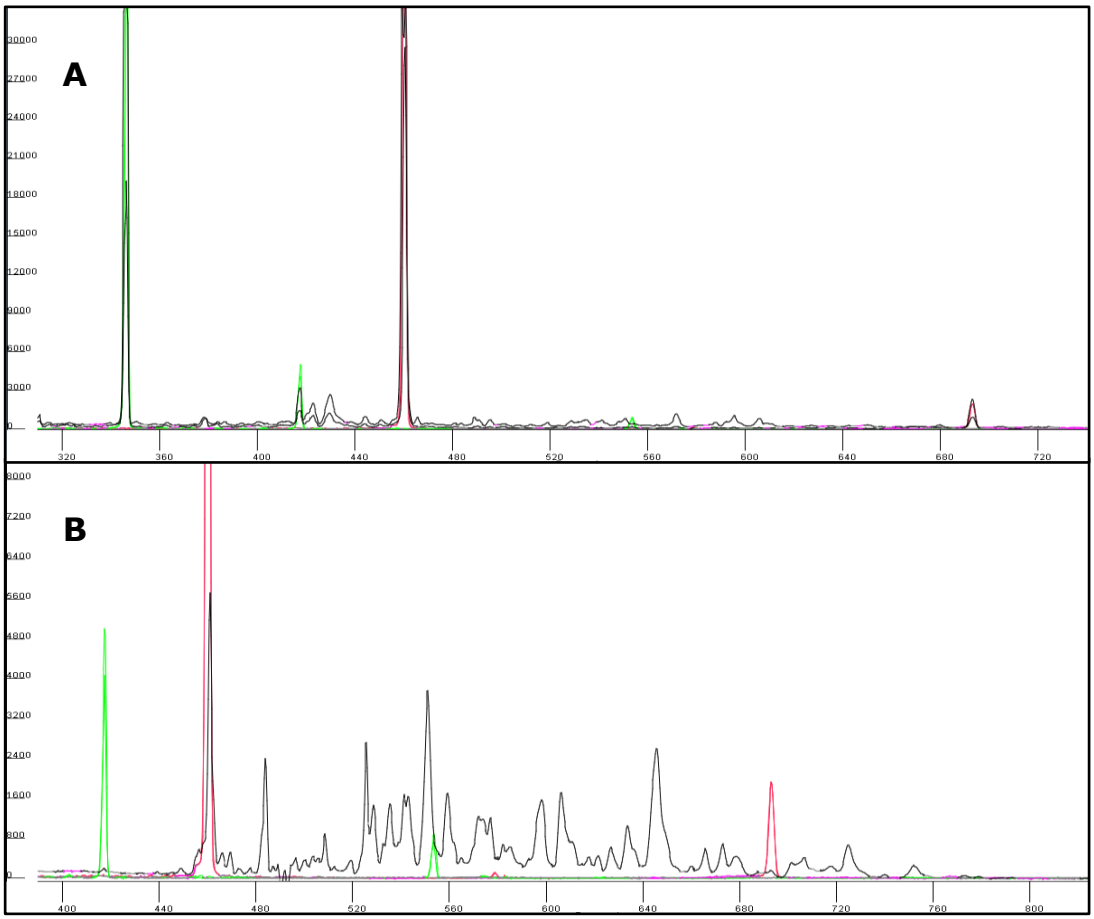


Figure 5.6 PCR amplification products of AMRISA protocol performed on lysate of 10-fold serial dilutions of *Pn. sordellii* cells. Lanes 2-7 are amplicons using primer set C, and 8 – 13 used 16S rDNA Clostridia primers. Lysate from cell dilutions ranging from 1.17 x 10^5 CFU in lanes 2 + 8 to 0.17 CFU in wells 7 + 13 used as templates. Bands were visible for the mPCR assay at 1.17 x 10^3 CFU, while amplifiable DNA was still detected at 1.17 CFU.



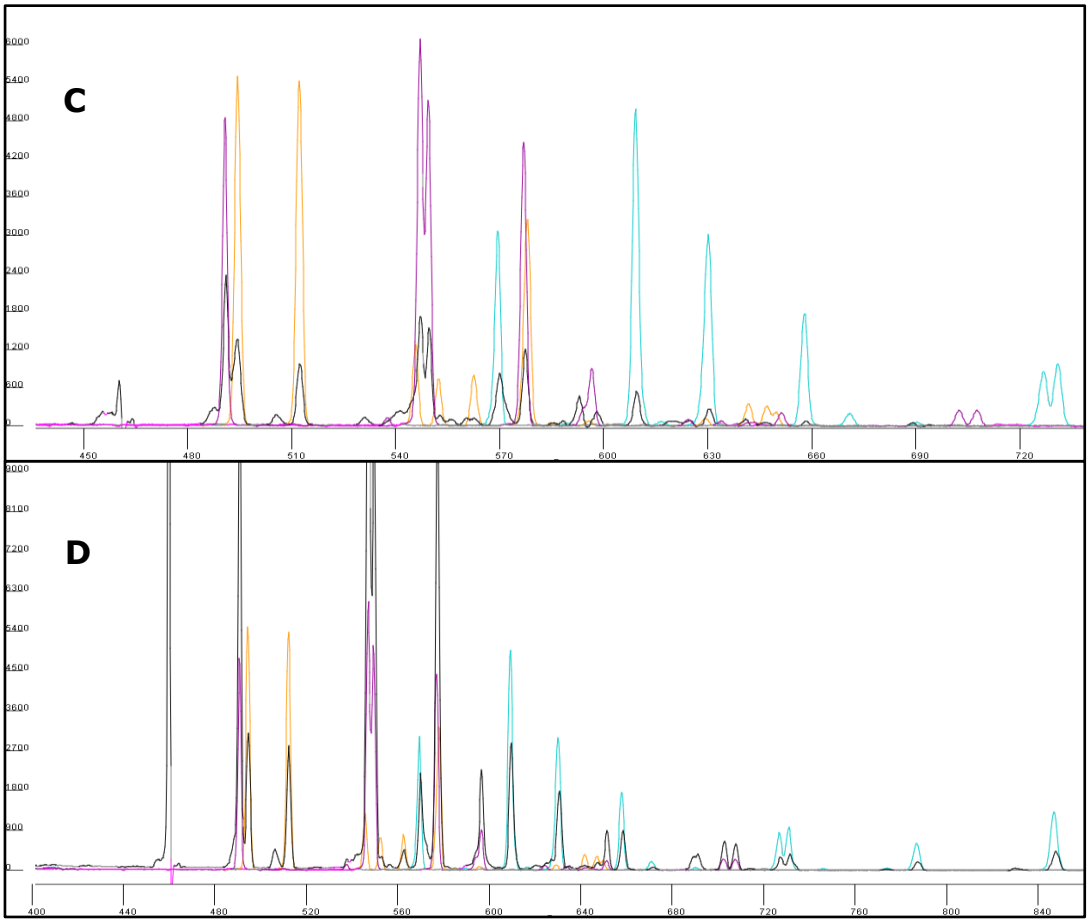


Figure 5.7 AMRISA profiles for environmental samples. Profiles (black line) for pathogen-spiked soil (A), agricultural soil (B), anaerobic digestate 1 (C) and anaerobic digestate 2 (D). Profiles for VIC channel only for A + B, and 6FAM channel only for C + D. The profiles for pure strains of *Cl. difficile* (teal), *C. perfringens* (red), *C. sporogenes* (green), *Pn. sordellii* (purple) and *Pr. bifermentans* (orange) have been overlaid.

Table 5.5 Summary of main detected peaks in environmental samples

(A) Spiked soil				(B) Farm soil			
Dye	**Peak**	**Peak height[1]**	**ID[2]**	**Dye**	**Peak**	**Peak height**	**ID**
6FAM	460	470		6FAM	491.7	10302	*PS*
6FAM	493	1664	*PB*	6FAM	531.3	2015	
6FAM	520	910		6FAM	545.8	1111	*PS*
6FAM	524	228		6FAM	576.5	928	*PS*
6FAM	529	260		6FAM	624.8	2704	
6FAM	546.5	190	*PS*	6FAM	630.4	744	*CD*
6FAM	571.7	230		VIC	345.7	1195	*SP*
6FAM	579	182		VIC	461	5493	*CP*
6FAM	603	125		VIC	484	2384	
6FAM	630.4	124	*CD*	VIC	525.6	2730	
VIC	346.3	19223	*SP*	VIC	541.3	1086	
VIC	417.4	1065	*SP*	VIC	543	1341	
VIC	422.9	1595		VIC	551	3584	
VIC	429.9	2413		VIC	560.3	1389	
VIC	460.4	30357	*CP*	VIC	598.2	1354	
VIC	571.8	1033		VIC	606.3	1556	
VIC	693.3	842	*CP*	VIC	645.8	2517	
(C) Digestate 1				**(D) Digestate 2**			
6FAM	490	370	PS	6FAM	493.3	3414	
6FAM	494.3	136	*PB*	6FAM	531.4	205	
6FAM	547	438	*PS*	6FAM	545.8	364	
6FAM	549.8	364	*PS*	6FAM	575.7	336	
6FAM	577	336	*PS*	6FAM	594.7	242	*PS*
6FAM	596.7	70	*PS*	6FAM	624.6	219	
6FAM	609	73	*CD*	6FAM	631.6	167	*CD*
VIC	460.5	9165	*CP*	VIC	460.4	390	*CP*
VIC	577.6	1539	*SP/SE*	VIC	483.5	4160	
VIC	627	333		VIC	490.9	4524	
VIC	651	370	*SP/SE*	VIC	556	1542	
VIC	692	667	*CP*	VIC	594	699	
				VIC	614	349	
				VIC	686.8	290	

[1] Peak height in relative fluorescence units (RFU)

[2] Match to pathogen peak: *Cl. difficile (CD), C. perfringens (CP), C. novyi (CN), C. septicum (SE), C. sporogenes (SP), Pn. sordellii (PS) and Pr. bifermentans (PB)*

5.5 DISCUSSION

Various gene targets and molecular approaches were investigated for their suitability for identifying and differentiating 16 important clostridial pathogens within 4 different clostridial genera. Until this point, molecular diagnostics have often favoured 16S rDNA-targeted approaches, although the highly conserved nature of this gene makes distinction between closely related clostridial species and strains unfeasible. In this study, a protocol with higher resolution capabilities was sought. Two different methods of discrimination were compared; using regions with a high degree of genetic polymorphism for restriction analysis following enzymatic digestion, or by amplifying regions which demonstrate natural length-heterogeneity (Marzorati *et al*., 2008). This was facilitated by T-RFLP or AMRISA methods: PCR-based techniques that utilise fluorescently-labelled primers for fragment-size analysis to identify particular microbial groups within samples (Marsh, 1999; Miyamoto, Kawahara and Minamisawa, 2004; Schütte *et al*., 2008; Hayashi *et al*., 2003; Cherif *et al*., 2008; Feligini *et al*., 2015). While these techniques are widely used to characterise complex microbial communities, a more targeted approach was enabled by designing primer sets selective to just the target pathogens, where possible. T-RFLP assays have been successfully used for simultaneous pathogen detection in other studies (Elliott *et al*., 2012; Nilsson and Strom, 2002), although not to identify clostridial pathogens to species level. Length-heterogeneity PCR has been successfully used to discern between similar *Clostridium* pathogens (Sasaki *et al.*, 2000; Song *et al.*, 2002), and even to ribotype different strains of the pathogenic *Cl. difficile* (Stubbs *et al*., 1999), yet a multispecies test capable of strain resolution has not yet been conceived. To facilitate the detection of multiple, genetically diverse species, while maintaining target specificity, multiplex-PCR was used for two protocols. This allowed for better primer T_m matching and reduced the chances of non-target amplification. The automated fragment analysis used with T-RFLP, ARISA and this protocol allows for more accurate sizing of fragments compared to gel electrophoresis-based methods such as

DGGE, and even enabled the simultaneous detection of pathogens in environmental matrices.

5.5.1 Evaluation of protocols

Three candidate assays were tested and compared utilising singleplex or multiplex PCR with degenerate or non-degenerate primers. PCRs with Primer Sets A and C performed as predicted *in silico,* whereas mPCR using Primer Set B had suboptimal reproducibility, spurious bands were frequently observed, and PCR conditions could not be optimised. It is postulated that very minor changes in mPCR conditions could bias amplification to one primer pair, or cause cross-reaction of degenerate primers to unintended targets, as described elsewhere (Henegariu *et al*., 1997; Pichon *et al*., 2006). PCR amplicons from Primer Set A were digested with *Sfc*I and *Ssp*I restriction enzymes before submission for FA. When visualised, fragment profiles were generally inconsistent with profiles predicted *in silico*. This may be due to suboptimal digestion conditions, or the over-addition of fluorescently labelled amplicons which could have induced star activity (lowered enzyme specificity for recognition sites) (Nagashima *et al*., 2003). Digests were rerun with amended conditions (shorter duration and lower amounts of amplicons) but this did not resolve the presence of spurious peaks after FA and profile visualisation. It is likely the discrepancy between predicted and observed peaks is due to genetic variation, incomplete digestion of amplicons or complex secondary structures produced during PCR (Elliott *et al*., 2012), although further optimisation could potentially resolve this issue. For the AMRISA protocol, a 1:75 dilution of amplicons from Primer Set C gave the best clarity after peak visualisation. Lower dilutions resulted in lower quality profiles and undesirable spectral pull-up in other dye signals. Higher dilutions often resulted in poor peak detection of smaller, identifying peaks. Post-PCR clean-up of DNA using ethanol precipitation prior to FA had little effect on peak profile clarity or interpretation, and therefore it was regarded as an unnecessary step. Alternate methods of PCR amplicon purification, such as

spin column techniques used with proprietary kits, could improve profile interpretability by removal of peak 'noise'. The lack of growth of *C. tetani* when cultured meant the efficacy of the PET-labelled *Cpath4r* primer could not be assessed. The PET channel was found to only detect *C. tetani* and *C. cochlearium in silico.* The absence of PET-peaks in environmental samples could therefore reflect either the absence of these bacteria from sample DNA, or the lack of primer sensitivity to targets. Peak size shifts of -3 to -4 bp for VIC and NED channels, and around -9 bp difference for 6FAM labelled fragments were observed. This difference between the sequence-determined (true) peak sizes and the observed size (drift), is frequently documented. Increased drift correlates to fragment molecular length, weight (possibly caused by increased purine content or dye label (Kaplan and Kitts, 2003)), G + C content and secondary structure melting point (Bukovská *et al*., 2010). It is possible that the increased drift in the 6FAM channel is due to the lower molecular weight of the 6FAM fluorophore (376.32) in comparison to the other dyes used (550). The true size of a fragment can be easily calculated, provided positive controls with known peak sizes are run with each round of samples to calculate drift.

5.5.2 Limit of detection, rapidity and applicability of assays

One of the aims of this study was to design an assay to identify low-level pathogen contamination which could be expected in some environmental reservoirs. While previous researchers have devised molecular arrays capable of discriminating between clostridial species of pure isolates (Janvilisri *et al*., 2010; Dhalluin *et al*., 2003; Song *et al*., 2002; Sasaki *et al*., 2002), the LOD and suitability for use on co-contaminated or complex environmental samples was not reported in these studies. Unlike other protocols, the use of a pre-enrichment step to aid the detection of low-level contamination was avoided (Lindström *et al*., 2001; De Medici *et al*., 2009), as this prolongs the turnaround time for detection and does not facilitate a quantitative investigation of the original level of pathogen contamination. The microarray assay of Janvilisri *et al.*, 2010 identified 4

pathogenic *Clostridium* species simultaneously from spiked horse stool samples, and could detect some individual pathogens at 10^4 CFU ml^{-1}. For Primer Set A, poor PCR yields, and unexpected/absent identifying peaks indicated the unsuitability of this assay for low-level pathogen detection and discrimination. The AMRISA protocol, with Primer Set C, demonstrated good LOD. Multiplexing is known to reduce detection sensitivity (Parker *et al*., 2015; Poritz *et al*., 2011), as primer-primer interaction products impede amplification. However, after 6-deep multiplexing, bands could still be visualised with gel electrophoresis using ~ 500 genome copies, or lysate from 1.17 x 10^3 culture-viable cells as templates for reactions. This LOD is a satisfactory compromise considering the number of pathogens detected yet allows the screening of samples with low-contamination levels. The LOD could be increased further by removing selected labelled primers if the full-target range is not required. Alternatively, the multiplexing could be performed in two parts; one mPCR with *Cpath*Bf, *Cpath*1r and *Cpath*2r, and one with *Cpath*Af, *Cpath*2r, *Cpath*3r and *Cpath*4r. Amplicons can then be pooled and prepared for FA as described. This would still have the same target scope, but LOD would likely be enhanced. Additionally, MPCRs were performed on gDNA recovered using simple Chelex or boiling-based DNA extraction methods. Target detection could be improved by purifying the gDNA, removing potential-PCR inhibitors and unwanted cellular debris. The higher detection sensitivity of the AMRISA-mPCR step, in comparison to the *pgk* methodologies, was attributed to the fact that the primer targets (rrn operons) occur at multiple sites in bacterial genomes (Song *et al*., 2002) providing more primer annealing sites per reaction. Interestingly, after FA and visualisation, peaks were observed for *C. novyi* despite not being visualised on gels. This demonstrates that when combined with an automated fragment analysis step, not only is fragment size calling accuracy improved, but the detection sensitivity of the overall protocol is enhanced. The AMRISA protocol allows a qualitative screen for pathogen contamination but could be developed to generate semi-quantitative data on contamination levels. Perceivably, an internal implication control (IAC) could be added to each sample at a range of concentrations, enabling the

quantification of target peaks, which could be back-calculated to equivalent pathogen concentration. Furthermore, in clostridial-community analysis studies, peak height (measured in relative fluorescence units) can be used to monitor changes in the relative abundance of peaks within the community.

With direct access to a sequencing machine, the methodology used for this assay allows 'sample to result' within a 24-hr timeframe. This rapidity is advantageous for healthcare practitioners when effective treatment is reliant upon the fast and accurate identification of pathogens, and in real-time investigations of food contamination. AMRISA was used to successfully detect *Cl. difficile, C. perfringens* and *C. sporogenes* at 10^4 – 10^6 CFU g^{-1} in spiked soils. *C. perfringens* was also identified in agricultural soil, along with the presence of other clostridial species. Anaerobic digestates contained multiple putative pathogens, including *Cl. difficile, C. perfringens, Pn. sordellii* and *Pr. bifermentans.* Clostridial pathogens have been isolated from digestates in other studies (Bagge, Sahlström and Albihn, 2005; Bonetta *et al*., 2014). This assay represents a suitable tool for monitoring clostridial pathogen presence in such samples, particularly when WGS is not amenable due to cost, time or data-processing limitations.

5.5.3 Future applications of AMRISA protocol

The utility of the AMRISA protocol was demonstrated on seven key clostridial pathogenic species. *In silico* analysis indicated that the protocol could distinguish between key pathotypes for several species. The assay could be a valuable tool for typing clinically important strains associated with disease outbreaks, such as for *Cl. difficile.* The protocol may facilitate direct strain identification from samples, without the need for isolating, or a pre-enrichment step first. The assay offers an inexpensive alternative to full-16S sequencing, and with the high-throughput nature of the test, its perceived main use is as a tool for environmental epidemiological studies. The multiplex strategy for detecting pathogens simultaneously offers

distinct cost and time advantages over singleplex approaches. The detection of other non-pathogenic species also facilitates the study of clostridial communities in samples. This could provide a useful method for identifying and exploring communities in environments where Clostridia are thought have integral functional roles, such as biogas digesters, gut and rumen microflora and cellulose-degrading communities. The assay could be deployed to investigate the abundance and interactions of clostridial pathogens and closely-related species in environmental reservoirs. Results can be visualised using free-to-use software, allowing fast and simple interpretation and pathogen identification with less subjectivity than other approaches. Additionally, molecular detection circumvents many issues associated with culture-based assays (non-culturable cells, atypical cell morphologies etc.).

5.6 CONCLUSIONS

Three candidate assays were designed and evaluated, all utilising PCR and downstream FA to detect and differentiate between key clostridial pathogens and toxinotypes. The AMRISA approach utilised natural length variation of the ITS region to distinguish between species and clinically important strains. Multiplexing facilitated greater specificity to the target pathogens which could be easily identified using peak visualisation software. In comparison, T-RFLP approaches targeting the *pgk* gene were less sensitive and specific. The LOD of the AMRISA assay was approximately 500 genomes per PCR reaction, which allows low level detection of cells present in natural environments. The utility of the assay was demonstrated on spiked and contaminated environmental samples, permitting simple, rapid identification of pathogens and other closely-related clostridial species, without the need for pre-enrichment of samples or DNA purification steps. The rapidity, cost-effectiveness and simple interpretation of the test provide flexibility for pathogen detection, typing and clostridial community analysis in clinical, food and environmental samples.

CHAPTER 6: CHARACTERISATION OF THE PRESENCE AND BEHAVIOUR OF CLOSTRIDIA IN SCOTTISH AGRICULTURAL SOILS

6.1 ABSTRACT

Livestock production is an important Scottish industry, with output worth around £1.7 billion in 2016 (Scottish Government, 2017). Clostridial disease presents a real and growing challenge to farmers, veterinarians and epidemiologists alike, with conditions such as blackleg, black disease, braxy and equine grass sickness causing significant herd losses and welfare issues. Many of these soil-based infections are known to be prevalent in Scotland, although the distribution and abundance of the causative pathogens in agricultural soils remains unclear. In this study, a clostridia-targeted automated multiplex ribosomal intergenic spacer analysis (AMRISA) was used to screen soils from four farms in north-east Scotland, each under different management regimes, for pathogens and other clostridial species. Soils were collected over a 12-month period (September 2016 - August 2017) from either persistently waterlogged (wet spots) or permanently well-drained (dry spots) sites. From soil DNA (*n=538*), 84 main peaks were identified, some of which were attributed to pathogens *Clostridium novyi* and *Paeniclostridium sordellii.* Peaks were also putatively associated with *C. baratii,* Group I *C. botulinum*, *C. chauvoei* and *C. perfringens.* Canonical correspondence analysis (CCA) and hurdle regression models were applied to ascertain the effect of farm type, season, wet/dry spot, soil water concentration and grazing intensity on changes in peak relative abundance. *Pn. sordellii* soil abundance was significantly higher at sites grazed by dairy cattle, in dry spots, and generally higher in pastoral vs. arable soils. *C. novyi* peaks were significantly higher in soils collected during colder, wetter months, and in fields where rotational sheep and cattle grazing occurred. Short-term fluctuations in relative soil water concentration did not affect abundance of

either pathogen. The epidemiological relevance of these results is discussed.

Keywords; Clostridia, *Clostridium,* Scotland, pathogen surveillance

6.2 INTRODUCTION

The soil is a major reservoir for clostridial pathogens. Soil contamination of wounds and foods as a main disease transmission route is well acknowledged, and researchers have been isolating clinically important pathogens such as *Clostridium botulinum, C. perfringens* and *C. tetani* from soil samples for over a century (Noble, 1915; Schoenholz and Meyer, 1922). However, the factors driving the spatial and temporal distribution of these pathogens in terms of prevalence in different locations and soil types, such as soil physicochemical characteristics, climate and human activities, remain poorly understood. Research into the prevalence and epidemiology of the debilitating livestock pathogens *C. chauvoei, C. septicum, C. novyi* and *Paeniclostridium sordellii* also remains limited. The role of agricultural soils as a pathogen reservoir is of particular interest as they represent an important source of infection for humans *via* the food chain, for livestock, and also are a large repository of antibiotic resistant pathogens and genes (Forsberg *et al*., 2012). Direct-wound contamination, ingestion of bacteria or spores (Wolf *et al*., 2017) or contamination of animal feed and produce (Songer, 1996; South, 2014) are all potential routes of infection for humans and animals. The longevity of the environmentally-persistent spores produced by most clostridia, combined with the regular practice of manure, slurry or anaerobic digestate incorporation into the soil could create elevated concentrations of these pathogens (Boschiroli *et al*., 2016; Notermans, Dufrenne and Oosterom, 1981; Pulvirenti *et al*., 2015). In Chapter 4, the potential for long-term persistence of clostridial pathogens after inoculation into soil was demonstrated in microcosm studies. Similarly, the persistence of *C. sporogenes, C. botulinum and C. perfringens* in soil has been demonstrated in field trials (Gessler and Böhnel, 2006; Brochier *et al*., 2012), as well as

the subsequent contamination of overlying crops (Girardin *et al*., 2005). These studies indicated that clostridial pathogens remain at consistent spore concentrations, apparently unaffected by seasonal or climatic influences (Gessler and Böhnel, 2006; Sandler *et al*., 1993; Girardin *et al*., 2005). This is however inconsistent with epidemiological data. Numerous studies have linked clostridial outbreaks to seasonal (Wolf *et al*., 2017; Wylie and Proudman, 2009; Lewis, 2007) and meteorological factors (Useh *et al*., 2006), indicating possible activation of bacteria or spores within the soil. Understanding the underlying causality of these outbreaks requires an improved understanding of baseline-soil pathogen levels, and their temporal behaviour.

High clostridial abundance has generally been associated with wet tropical soils (Pett-Ridge and Firestone, 2005), rice paddies (Chin *et al*., 1998; Weber, Stubner and Conrad, 2001; Liu *et al*., 2009) and wetland sediments (Sandler *et al*., 1993; Espelund and Klaveness, 2014) owing, in part, to the anoxic nature of these environments (Hengstmann *et al*., 1999; Grimes, 1991). Frequent or long-term waterlogging in temperate, agricultural soils may also provide anoxic soil conditions capable of supporting clostridial pathogen growth. *C. tetanus* and *C. chauvoei* detection rates were significantly higher after flooding events in Taiwanese soils (Huang *et al.*, 2013) and in Zambia, most bovine clostridial cases between 1985-1994 were reported from cattle grazed on the plains of the Zambezi river (Munang *et al*., 1996). Cases of fatal blackleg in Styria, Austria, were higher in areas with underlying impermeable rock (Wolf *et al*., 2017). These findings indicate that soil waterlogging or raised soil water concentrations could affect causative microbes and/or susceptibility of hosts. When indigenous or introduced pathogens (via faecal shedding from livestock, wildlife, or from organic matter incorporation) are already present in soils, growth may be sustained or enhanced, creating areas or hotspots of elevated contamination. Spores may become activated into a virulent, vegetative form in wet soils, with high rainfall thought to be a predisposing factor for blackleg outbreaks in Nigeria (Useh *et al*., 2006;

Useh, Nok and Esievo, 2006). In contrast, dry soils can also sustain elevated pathogen concentrations. In Chapter 4, higher pathogens levels were observed for >84 days in drier soils (25% WHC) than in wetter soils, potentially attributable to the higher and quicker rate of transition of cells into spore form. One case-control study of equine-grass sickness (EGS) found that 66% of case occurred after 2-week periods of predominately dry weather (Wood, Milne and Doxey, 1998).

The soil mineralogy, structure and drainage can all influence the indigenous microbial communities, as discussed in Chapter 2. The adsorption rates of bacteria to soil particles is highly dependent on soil type and the presence of organic material (Mosaddeghi *et al*., 2009). Differences in soil porosity alter the number of micro-niches suitable for pathogen growth. Furthermore, the pore structure can determine the level of pathogen retention/carriage through the soil (Stoddard, Coyne and Grove, 1998; Mosaddeghi *et al*., 2009). The translocation of *C. botulinum* spores from upper to lower soil horizons has been demonstrated (Gessler and Böhnel, 2006). While this could reduce the pathogen load in the overlying soil layers, it could in turn result in contamination of groundwater (Stoddard, Coyne and Grove, 1998). Where movement of pore water is limited, such as in water-logged soils, Clostridia may persist.

Previous research also indicates that the type and intensity of farming practices may affect clostridial behaviour and abundance in the soil (Voidarou *et al*., 2011). Addition of faecal matter has been shown to be a source of soil *C. perfringens* and *C. septicum* contamination (Lewis, 2007; Voidarou *et al*., 2011). The addition of organic matter to soil has also been shown to enhance clostridial proliferation (Garcia and McKay, 1969). Relationships between grazing type, density and duration and the level of soil contamination are not clear. *Clostridioides difficile* was isolated from 37% of soil samples collected from communal grazing areas in rural Zimbabwe, with chicken faecal matter the likely source of contamination (Simango, 2006). Soil disturbances have been implicated with increased

risk of blackleg and EGS (Wylie and Proudman, 2009), and consequently the type and frequency of soil-disturbing practices, such as tillage or any digging of draining, could affect the soil clostridial community. While tentative links have been made between land management practices and the onset of clostridial diseases (Mccarthy *et al*., 2010; Newton *et al*., 2010), the prevalence of most clostridial pathogens in the soil and the factors driving their distribution, behaviour and abundance have not yet been elucidated.

In this study, a novel, high-throughput molecular diagnostic tool was employed to identify and explore the behaviour of clostridial pathogens and communities in agricultural soils across NE Scotland. Crop-damaging *Clostridium* species have been previously identified in nearby soils despite cool annual temperatures (Perry, 1982, 1985). Scotland also appears to have an association with braxy in sheep (Lewis, 2011, 2007), equine grass sickness (Wylie *et al*., 2014) and has recently experienced increased reports of blackleg (Scottish Agricultral College, 2012). The climate, soil type and importance of livestock-based agricultural in Scotland accentuate the need for research into clostridial-disease, which in turn necessitates a better understanding of the prevalence and causes of clostridial soil contamination in the region. It was hypothesised that soil, management and meteorological factors cause temporal changes in clostridial pathogen presence and community change. The key hypotheses were that clostridial pathogen presence and abundance are affected by i) changes in soil water concentration, ii) farm management and iii) changing environmental conditions. Changes in the overall clostridial community were also explored.

6.3 MATERIALS AND METHODS

6.3.1 Farm and site selection

Agricultural soils were collected from twelve sites in total across four farms in NE Scotland, along the 'A90 corridor' (Figure 6.1). Farms 1 and 2 are the northernmost farms, located in Aberdeenshire, UK, and Farms 3 and 4 were situated outside Brechin, Angus. The farms were selected based on their different land management regimes (Table 6.1). The dominant regimes at each farm were broadly categorised as Farm 1: mixed grazing (MG) where both beef cattle and sheep are grazed at various points of the year; Farm 2: beef finishing, (B) where beef cows and followers are grazed outdoors for most the year; Farm 3: dairy (D), where dairy cattle are grazed for short periods over the summer; Farm 4: an arable (A) catchment used solely for grain production. Approximate herd sizes and breeds are given in Table 6.1. It was assumed that all livestock faecal inputs, whether applied *via* manure/slurry spreading or directly from grazing animals, will act as an input of clostridial pathogens.

Three sampling locations were chosen at each farm. Two sampling sites were identified as potential "wet hotspots". These were localised areas of soil that remain naturally wet or waterlogged throughout the entirety of the year, likely having the anoxic soil conditions hypothesised to support the growth of anaerobes. Suitable waterlogged areas were pre-emptively identified using the James Hutton Institute soil data visualised in ArcGIS (ArcMap, ed. 10.2.1. ERSI, USA); namely topological recesses in areas overlying imperfectly or poorly-drained soils. Areas were then surveyed in late summer to identify exact hotspots. Sites where cases of clostridial disease had been reported where selected preferentially. Additionally, one "dry" site was selected at each farm, based on the prerequisite conditions of having well-drained soil and/or in a comparatively elevated position while spatially close (500m) to a wet hotspot (Figure 6.1). The soil series and drainage class for each wet/dry sampling site are given in Table 6.1.

6.3.2 Sample collection

Transects were laid from the centre of wet or "dry" spots towards comparatively drier (for wet spots) or wetter (for dry spots) soil regions, with the aim of maximising the water gradient across the transect. In addition, a transect was taken away from the "dry" spots towards an area likely to have a greater soil moisture. Soil cores were collected from within a 30 cm^2 area, in triplicate, at distances of 0, 4 and 16 m along the transect from the wet/dry spot. Cores were obtained by hammering an ethanol-sterilised 60 mm x 12 mm cold drawn steel tube 30 mm into the soil and expelling the core with a clean steel 10 mm rod into a sterile 7 ml polypropene Bijou bottle. Overlying vegetation (if present) was gently removed using a clean trowel to expose topsoil. Care was taken to avoid sampling on, or directly next to, livestock or wildlife faecal matter. Soil samples were immediately stored and transferred at 4°C, snap-frozen in liquid nitrogen within 24 hours of sampling and stored at -20°C. Soil samples were taken from proximal points every 6 weeks for 54 weeks (10 sampling events). Each sampling event was completed over a two-day period. The grazing regime was recorded for each site at each sampling date, with '0' representing no-grazing or no organic soil amendment (OSA) within the last 6 months, '1' indicating very light grazing or grazing/OSA at site <6 but >3 months previously, '2'; grazed or OSA added within last 3 months but not currently, and '3' indicated that site was currently being grazed or had recently had OSAs. Meteorological data were also obtained from the James Hutton Institute automatic weather station located at Farm 1. This included average daily air and soil temperature (10 cm depth) and average precipitation (mm) data.

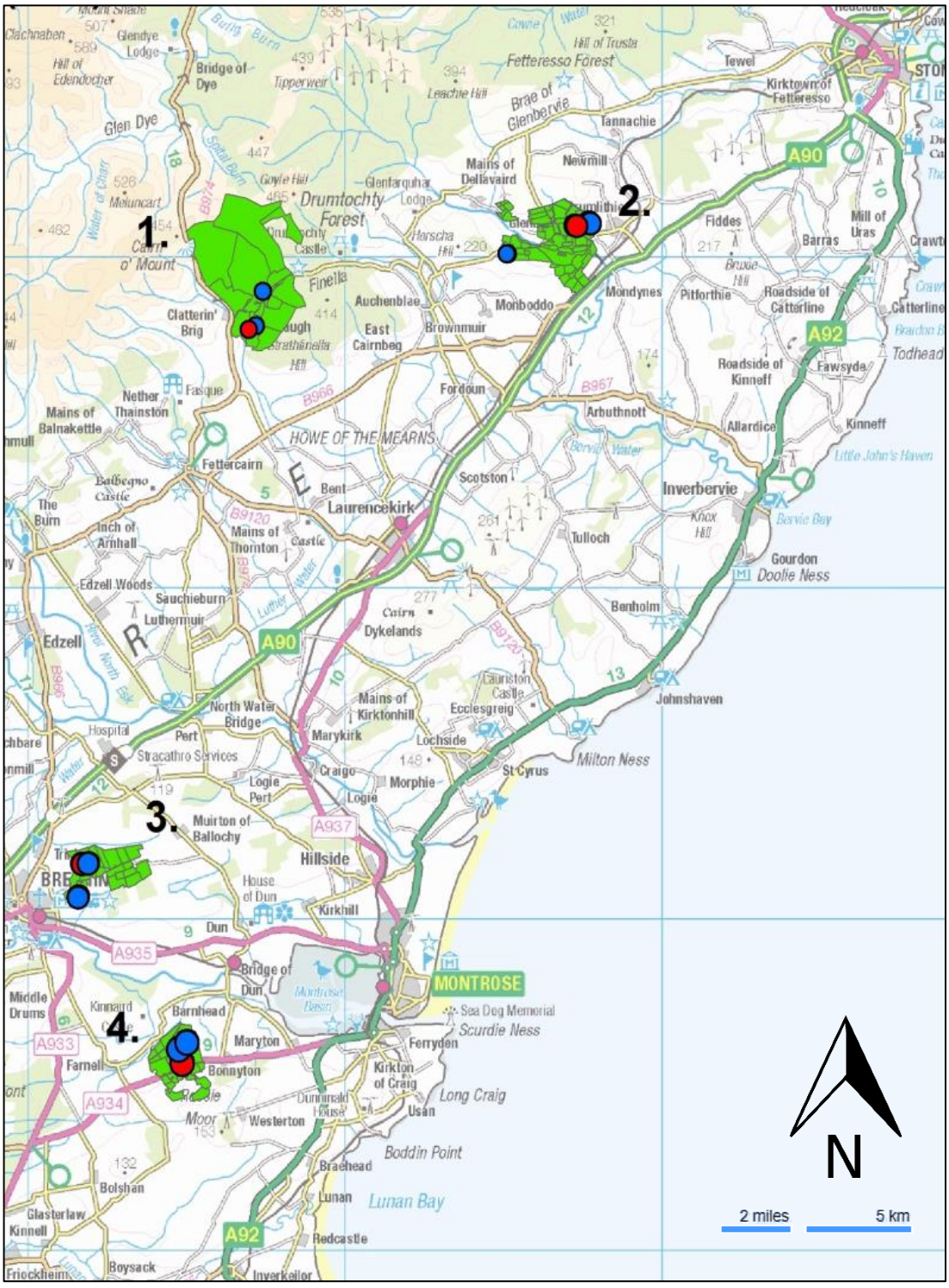


Figure 6.1 Location of the four study farms in NE Scotland. The approximate position of wet hotspots or dryspot sampling sites at each farm are denoted by blue and red dots, respectively.

6.3.3 Soil moisture and relative moisture

Volumetric soil moisture, θv, which is the ratio between the volume of water present and the total volume of a soil sample, was determined using a non-destructive ThetaProbe Type ML2x moisture probe (Delta-T, UK) with a Delta-T HH2 data logger (Delta-T, UK). Soil moisture readings were used to indicate the soil moisture across sampling transects throughout the year. Readings were collected in triplicate at each sampling point. Pre-set meter calibrations for either mineral or organic soils were used depending on the predetermined organic content of the soil at sampling locations, whereby <7% = mineral soil and >7% organic soil. As θv varied depending on soil type and bulk density, the relative soil moisture was calculated by deducting the site average θv over the entire sampling period from each individual θv value.

6.3.4 Soil DNA extraction

DNA was extracted from soil samples using the PowerSoil® DNA 96-well Isolation Kit (MoBio Laboratories, CA, USA). On the day of extraction, soils were thawed on ice and 250 mg were added aseptically to an individual well on the extraction plate. DNA was extracted following the manufacturer's instructions, except with the inclusion of 2 freeze/thaw cycles (stored at -70°C until frozen then immediately thawed in a 65°C water bath) after the addition of the PowerSoil Bead Solution, followed by heating to 70°C for 15 min prior to bead beating. These additional steps were recommended by the manufacturer to help improve the lysing efficiency from bacterial spores. A positive control soil was run for each extraction plate. This was a soil which had been spiked with *Cl. difficile, C. perfringens* and *C. sporogenes* (see Chapter 3 for details). To check for successful DNA extraction, the extract was visualised using gel electrophoresis in a 1% (w/v) agarose gel strained with ethidium bromide. To assess the quantity and purity of DNA, extracts were also assessed by spectrophotometry using a Qubit® 2.0 Fluorometer

Table 6.1 Farm and sampling site management and soil characteristics

Farm	Farm size (crop)	Site code[1]	Soil series and MSSG[2]	Drainage Class[3]
1	1,000 ha (990 ewes + 50 suckler cows)	MG.H1	Corby, Humus-iron podzol	Freely drained
		MG.H2	Anniegathel, Noncalcarous gley	Poorly drained
		MG.C1	Balkies, Brown earths	Imperfectly drained
2	445 ha (450 Aberdeen Angus cows + calves)	B.H1	Barras, Noncalcarous gley	Poorly drained
		B.H2	Shields, Humus-iron podzol	Undifferentiated class
		B.C1	Shields, Humus-iron podzol	Freely drained
3	121 ha (330 Holstein cows + calves)	D.H1	Balrownie, Brown earths	Imperfectly drained
		D.H2	Balrownie, Brown earths	Imperfectly drained
		D.C1	Balrownie, Brown earths	Imperfectly drained
4	284 ha	A.H1	Pow, Noncalcarous gley	Poorly drained
		A.H2	Pow, Noncalcarous gley	Poorly drained
		A.C1	Vinny, Humus-iron podzol	Freely drained

[1] Site codes denote mixed grazing (MG), beef grazing (B), dairy grazing (D) and arable (A) management types at hotspot (H) or dryspot (C) sampling sites. [2] Soil series and MSSG (major sub-soil group) based on Scottish soil classification https://www.hutton.ac.uk/learning/soilshutton/soil-classification. [3] Drainage class based on the Hydrology of soil types (HOST) classification (Boorman, Hollis and Lilly, 1995)

(Invitrogen, Paisley, UK) with the Qubit Broad Range assay or with the NanoDrop 2000c spectrophotometer (ThermoFisher, UK). For both techniques, 1 µl of DNA was used to calculate DNA concentration and the absorbance ratio at both 260/280 and 230/260 nm wavelengths.

6.3.5 Clostridia targeted AMRISA

DNA from successfully extracted samples was screened for the presence of clostridial pathogens using the automated multiplex ribosomal intergenic spacer analysis (AMRISA) protocol with Primer set 3, designed and optimised in Chapter 5. *In silico* PCR analysis with this primer set in Chapter 5 identified high specificity to most clostridial pathogens and other important clostridial species, with the natural length heterogeneity of amplicons allowing discrimination between target species and strains. Full AMRISA was then performed on 7 test species, verifying the sensitivity and specificity of the protocol. In this experiment, the AMRISA protocol was split into two separate multiplex polymerase chain reactions (mPCR) to further increase the sensitivity of the method. The amplicons from both mPCRs were then combined before fragment analysis (FA). The first primer set (A) consists of forward primer *CpathAf* and three fluorescently-labelled reverse primers, *Cpath2r, Cpath3r* and *Cpath4r* (Table 6.2). The second primer set (B) consists of forward primer *CpathBf* and two fluorescently-labelled reverse primers, *Cpath1r* and *Cpath2r* (Table 6.2; see section 5.4.2 for primer sequences, targets and T_m).

The separate mPCRs were performed on each soil DNA extract. Reactions were performed in sterile, polypropylene 96-well PCR plates (ThermoFisher Scientific, UK). In each well, a 25 µl PCR reaction mixture containing: 50 ng of soil DNA template, 0.75 µl (20mM) of each primer from set A or B, 2.5 µl of 10 × NH_4 buffer, 2 µl of 50mM $MgCl_2$, 1 µl of 20 mM dNTP mix, 0.5 µl of BSA and 1.25 U of *BioTaq* DNA polymerase was prepared. No-template controls (NTC) were prepared with sterile molecular water and run with each mPCR plate, in addition to DNA extracted from the positive control soil. Thermal cycling was performed

Table 6.2 Primers used in this study and their targets

Set	Forward	Reverse	Dye label	Target pathogens
A	*CpathAf*	*Cpath2r*	**VIC**	*C. botulinum* Group I + *C. sporogenes*
		Cpath3r	**NED**	*C. botulinum* Group III, *C. haemolyticum* + *C. novyi*
		Cpath4r	**PET**	*C. tetani*
B	*CpathBf*	*Cpath1r*	**6FAM**	*Cl. difficile, Pn. sordellii* + *Pr. bifermentans*
		Cpath2r	**VIC**	*C. baratii, C. butyricum, C. chauvoei, C. botulinum* Group II, *C. perfringens* + *C. septicum*

as follows: 95°C for 3 min followed by 32 cycles of 94°C for 30 sec, 56°C for 30 sec and 72°C for 52

6.3.6 Fragment Analysis (FA)

For each PCR-positive sample, 1 µl of amplicon from each primer set was diluted 1:75 in SDIW and 2 µl combined with 11.8 µl HiDi® Formamide (ThermoFisher Scientific, UK) and 0.2 µl of GeneScan 1200 LIZ internal size standard (Applied Biosystems, UK), before submission for FA (ABI PRISM 3730 Genetic Analyzer (Applied Biosystems, UK), the James Hutton Institute, Dundee, UK).

Raw FA profiles were imported into GeneMapper™ software v4.0 (Applied Biosystems, UK) for peak detection and size calling. A modification of the 'AFLP default' analysis method was used; peaks between 250-1200 base pairs (bp) and above 50 relative fluorescence units (RFU) were sized using the Local Southern Method, and the analysis performed. Peaks outside this size range are likely to be PCR artefacts (such as primer-dimers), while peaks less than 50 RFU are typically spectral-pull up, neither being representative of actual target amplicons. Profiles with sizing quality values <0.1 were excluded from the analysis. A comma separated file containing columns for sample name, dye, peak size (bp), peak height (RFU) and peak area data for each peak within a profile was exported from GeneMapper and imported into *R*. Functions within the draft library TRFLPR (https://github.com/petersensm/TRFLPR) were utilised within *R* to bin, normalise and remove 'noise' from the peak data prior to further analysis. Firstly, functions from the reshape2 and plyr libraries (Wickham, 2007; Wickham and Wickham, 2018) were used to 'clean' and process data into a data frame for each dye. Subsequent analyses from this point were performed separately for each dye, due to the differences in detection sensitivity, background noise levels and binning thresholds between dyes. Basic *R* functions were used to exclude unsized, 'tiny' (<250 bp) and 'large' (>1200 bp) fragments. Peak noise thresholds (in RFU) were determined separately for each dye, and fragments below this value were excluded. The relative abundance obtained for each peak within a sample, was used *in lieu* of the raw peak height or area. This was calculated by dividing a peak's height by the sum of the heights of all peaks within a profile. Peak heights were used instead of peak areas to calculate relative abundance as this has been found to give better signal to noise ratios (Culman *et al*., 2008). Rare fragments with relative abundances <1% were omitted, and the relative abundance of remaining peaks recalculated.

Peaks were then binned in two stages. Firstly, fragments separated by a minimum of 0.75 bp were clustered into bins using the bincimate function of TRFLPR. Next, bins with an inter-bin difference of <1.5 bp between

averages bin sizes were consolidated together using the lump.bins function. Each bin was labelled with the average size (bp) using the tagBins function, and summary matrix of binned peaks and their relative abundance was created using reshape2's dcast function. Data frames for each dye were merged, and samples with no peak data were removed from the data frame. Quality summaries were produced to determine the number of peaks discarded from the analysis for each dye.

6.3.7 Identification of AMRISA peaks

Binned peaks were putatively identified by cross-referencing with the peak library generated in Chapter 5 (Supplementary Table 3). Small peak size shifts between observed and predicted peak lengths (drift) are frequently observed due to inaccuracies in fragment measurement by the sequencing machine, namely due to different characteristics of the fragments (Kaplan and Kitts, 2003; Bukovská *et al*., 2010). To calculate drift, AMRISA was performed with seven clostridial strains with known fragment lengths. These were run simultaneously with the soil samples, and a drift correction factor was derived based on the difference between observed and predicted peak sizes. These positive samples were also used to help identify pathogens from soil samples. When peaks could not be associated with a species, peaks that clustered together on canonical correspondence analysis (CCA) ordination plots (see Section 6.3.8.2) or positively correlated in terms of their relative abundance were tentatively assumed to belong to the same species.

6.3.8 Statistical analysis

6.3.8.1 Clostridial population metrics

The total number of unique taxa (S), Shannon index (*H'*) (Equation 1) and Pielou's measure of species evenness (*J'*) (Equation 2) were calculated for each profile after data normalization, noise removal and peak binning, under the assumption that each peak represented a unique taxon. These

indices were calculated in *R* using the diversity and specnumber functions of the vegan package (Oksanen *et al*., 2013)**.**

$$H' = -\sum_{i=1}^{S} p_i \ \mathrm{In}\, p_i \qquad (1)$$

Whereby p_i is the number of individuals of a taxa, divided by the total number of samples.

$$J' = \frac{H\prime}{\log(S)} \qquad (2)$$

6.3.8.2 Correlations between peaks

Spearman's rank correlation coefficients were calculated between the relative abundance data for each main peak with a dye set. Two-sided significance tests were performed to assess the significance of the correlation, with p > 0.05 deemed significant. Correlation matrixes were produced using the rcorr function from the Hmisc library (Harrell Jr and Harrell Jr, 2018).

6.3.9 Canonical correspondence analysis

Canonical correspondence analysis was used to explore changes in community composition of targeted Clostridia with changes in environmental and management variables. Canonical correspondence analysis is a distance-based multivariate analysis technique allowing the ordination and visualisation of multi-species abundance data in relation to environmental variation (Ter Braak, 1986, 1987). The technique is used when a unimodal relationship between a microbial community and environmental gradient is expected, rather than a linear relationship (Schütte *et al*., 2008; ter Braak and Šmilauer, 2015). The method is shown to be advantageous in ecological studies compared to other distanced-based relatives such as non-metric multidimensional scaling (NMDS) (ter Braak and Šmilauer, 2015). Weighted averaging methods such as CCA

suppress the effect of rare species on the overall analysis, and are not hampered by zero inflation caused by numerous zeroes in species data (ter Braak and Šmilauer, 2015). Canonical correspondence analysis was performed using the vegan library in *R* to assess the effect of relative soil water concentration, sampling date, site type and grazing level on the relative abundance of targeted clostridial species across the four different farm types. Furthermore, close clustering of peaks on ordination plots was used to indicate fragments belonging to an individual species. A data frame containing relative abundance data for main binned peaks (present in >10% of samples), for all dyes was initially analysed against all environmental predictors (including farm type) using the vegan's cca function. Variance inflation factors (VIF) were determined using the vif.caa function in vegan to check for multicollinearity between predictors. If any predictor had a VIF over 10, the variable with the highest VIF was dropped from the model and the CCA repeated, minus the redundant variable, until all VIF values were under 10. Analysis of variance (ANOVA)-like permutation tests were performed using vegan's anova.cca function to assess the significance of the model, the model terms (environmental variables) and each constrained axis, respectively. Tests were run with 999 permutations, with $p < 0.05$ indicating a significant model or component. A partial-CCA was also performed as described above, with the exception that the effect of farm type is removed before analysing the effects of the other environmental variables on the dataset. Ordination plots were produced for both models to graphically interpret the results of the CCA.

6.3.10 Peak and diversity response to environmental variables

A generalised linear model (GLM) procedure was performed in *R* to assess the response of total peaks (S) to explanatory variables. Post hoc pairwise comparisons were performed for main affects using Tukey's HSD with the mcp function from the library multicomp (Reeves, Farnell and Lan, 1994).For the population metrics *H'* and *J'*, non-parametric Kruskal-Wallis tests with multiple comparisons were used to assess difference in diversity

between farms, harvest date and wet spots vs. dry spots. Multiple comparisons were made using Dunn tests, performed with the dunnTest function in the FSA package (Ogle, 2015) in *R.* Negative-binomial (NB) or hurdle models were used, where appropriate, to evaluate the effect of predictors on presence and the relative abundance of selected peaks. Hurdle models were used to compensate for high zero-counts and overdispersion in the response data (Potts and Elith, 2006; Zeileis, Kleiber and Jackman, 2008). These models are two-component models, which first model the presence-absence of a peak using a GLM with logit-link and negative binomial error, then model the abundance of a peak, where they were present (Heinänen, Rönkä and Von Numers, 2008). This modelling approach can help identify whether the presence/absence and the positive abundance response are driven by different predictors. An iterative approach was used to select the correct family distribution for all models used, whereby the appropriate model was selected based on Log-Likelihood and AIC scores, plus the closeness between observed and predicted zero-counts for the model. Repeated measures tests were not necessary because while samples were collected at the same locations along transects, gaps of at least 15 cm were left between previous coring points. The full-model was initially fitted, then interaction and main terms were dropped sequentially until the minimum-adequate model was produced. The effect of dropping a term was assessed using Wald tests (from library lmtest in *R* (Hothorn *et al*., 2015)). NB models were implemented using the glm.nb function of library MASS (Venables and Ripley, 2002) while hurdle models used the hurdle function from library pscl (Jackman *et al*., 2007).

6.4 RESULTS

6.4.1 Soil collection and processing

Soil samples were collected, along with grazing and other land-management data at every sampling period. Daily air and soil temperature and precipitation data were obtained for the entire sampling period from

the automatic weather station at Farm 1 (Figure 6.2), and soil volumetric water (θv) was measured at every sampling site (except for on the last sampling date due to a breakdown of the moisture probe). Soil volumetric water was used to calculate the relative water concentration for each sampling point. DNA was extracted from soils from every other sampling event ($n = 540$), corresponding to harvests in September and November 2016, February, May and August 2017. Gel visualisation confirmed that DNA extraction had failed for 2 soils samples. The AMRISA was performed on DNA extracts, and amplification products were electrophoresed on agarose gels to determine mPCR success. Bands for PCR amplification products were present for 533 soil samples (98.7%) for at least one of the primer sets. Contamination was detected in 1 of 12 NTC reactions, most likely due to well-to-well transfer between adjacent samples. Amplicons for both primer sets were pooled, diluted 1:75 in SDIW, and 2 µl submitted for FA. Peak profiles were sized for 482 samples, including 6 NTCs. The remaining samples could not be automatically sized due to poor sizing-quality. It was also not possible to manually call size-standard peaks using the GeneMapper interface.

The peak quality summary for each dye is given in Table 6.3. NTCs contained no detectable peaks above the set thresholds and were excluded from the dataset, along with blank profiles, resulting in 475 remaining profiles. Profile analysis of 7 test clostridial spp. showed consistent peak drift of -8 to -9 bp for 6FAM, -3 to -4 bp for VIC, and -5 to -6 bp for NED channels respectively. Drift was not determined for the PET channel.

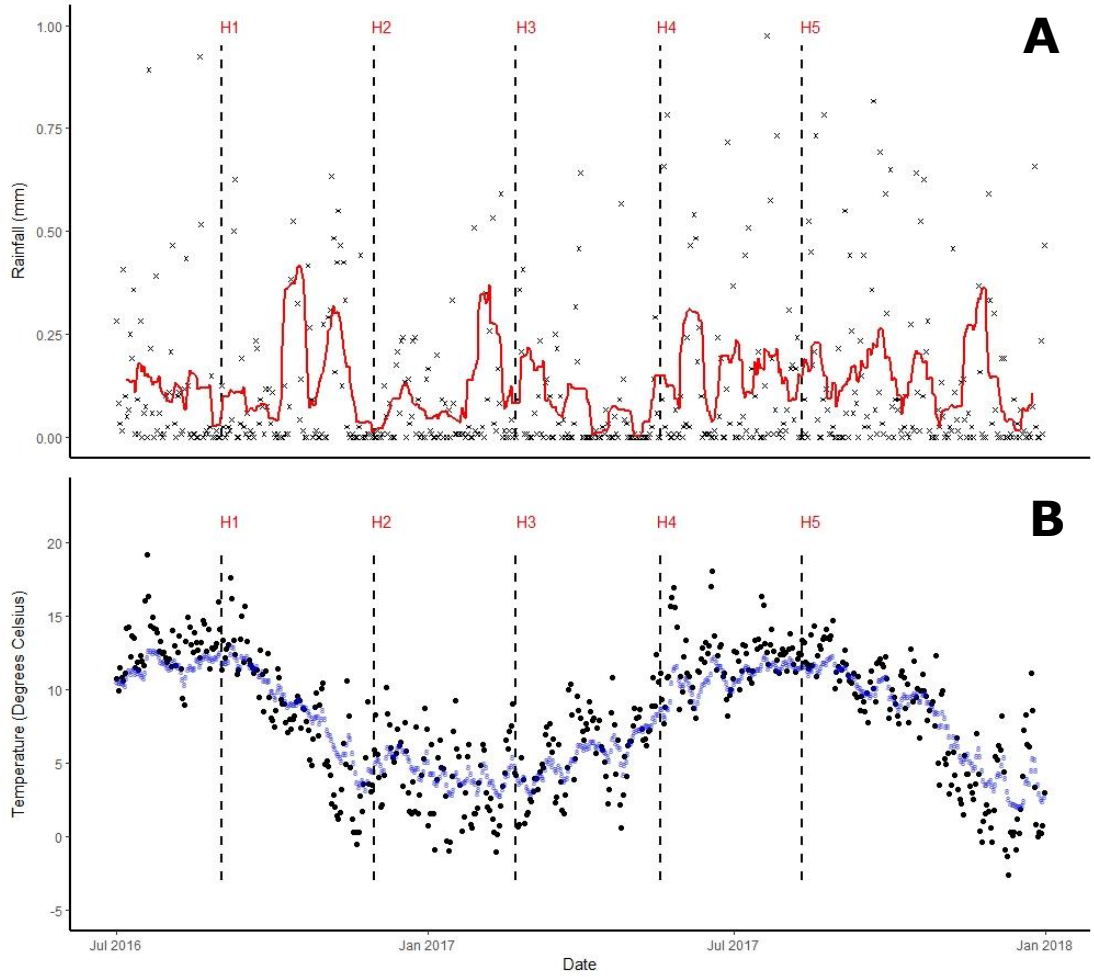


Figure 6.2 Meteorological data recorded from automatic weather station at Farm 1 for the duration of the sampling period. Panel A: Average daily rainfall (mm) values, with 2 week rolling average rainfall depicted by the red line. Panel B: Average daily air temperature readings (black dots) and the recorded soil temperature at 10cm depth (blue dots), both in degrees Celsius. Soil collection and DNA extraction events (H) are indicated with dotted lines.

6.4.2 Clostridial population metrics

The total number of peaks observed was significantly lower in February than at any other sampling time (p =<0.001; see Figure 6.3). The number of unique clostridial peaks (*S)* was also lower in May than in August (z= -3.00, p = 0.025). Dairy and MG regimes had lower total peaks than arable sites (z=-2.87, p=0.021 and z=-3.40, p=0.004, respectively).

Table 6.3 Peak quality control summary

Dye	Good	Threshold[1]	Noise	Tiny area[2]	ALL
6FAM	5399	0	0	276	5675
VIC	9581	100	23962	2885	36428
NED	5241	100	97859	67	103167
PET	800	150	94183	1	94984

[1] Noise threshold in RFU, below which peaks were discarded

[2] Peaks with a relative abundance of <1 % were omitted

Shannon Index (*H'*) significantly differed between harvesting events (χ^2 = 137, p = <0.001), with *H'* lower in February samples than all other months (p = <0.001; see Figure 6.4), and higher in August than in November, February and May (p = <0.05). Farm type had no significant effect on *H'* (χ^2 = 5.43, p = 0.142), although *H'* was significantly higher in permanently wet sites than dry sites (χ^2 = 5.12, p = 0.02).

Pielou's evenness of peaks (*J')* varied slightly between samples (Figure 6.5). Significant changes in *J'* were observed between sampling date (χ^2 = 18.5, p < 0.001) and between farm types (χ^2 = 23.1, p < 0.001). observed *J'* was significantly lower in May than in August (z= 3.61, p = 0.003), February and September (z= 2.82, p = 0.016 and z= -3.57, p= 0.002, respectively). Additionally, *J'* was significantly lower in soils under dairy management compared to arable (z= 4.23, p < 0.001) and MG (z= -3.82, p < 0.001) regimes, and *J'* was also comparatively higher in wet hotspots than in dry spots (χ^2 = 8.79, p = 0.003).

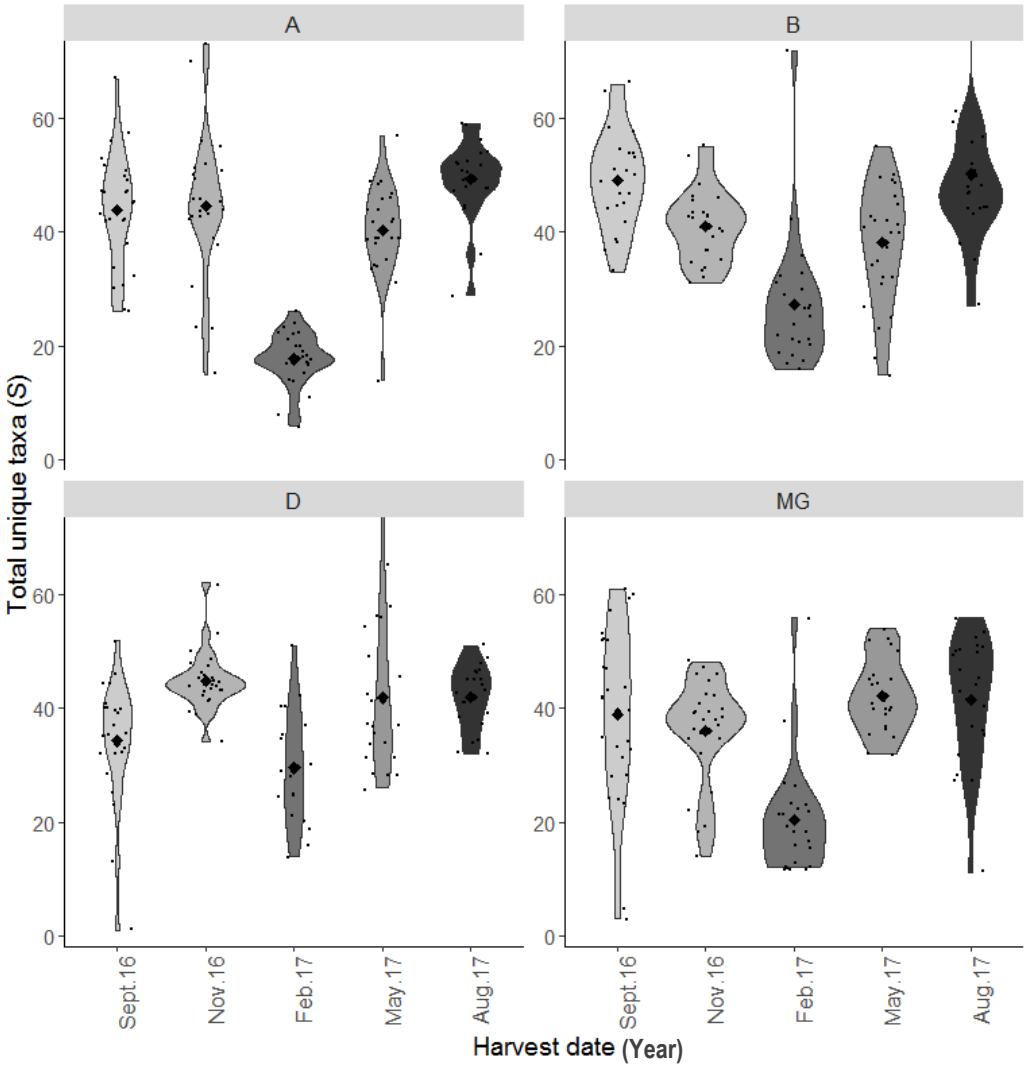


Figure 6.3 Trimmed violin plots showing the total number of unique peaks (*S*) across sampling regime at different farms. Arable (A), beef (B), dairy (D) and mixed grazing (MG). Shaded area displays the kernel probability density. Large black dot indicates mean total peak number for each group.

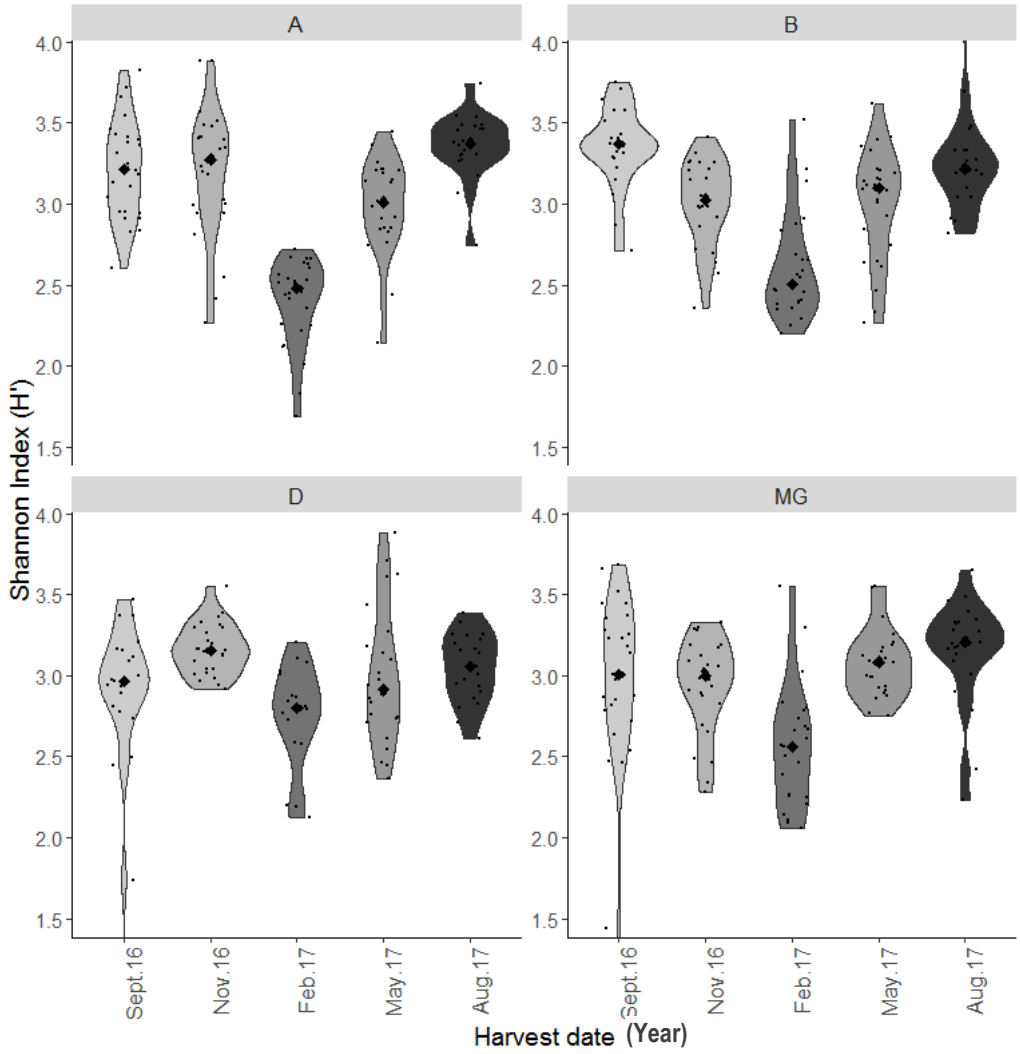


Figure 6.4 Trimmed violin plots showing Shannon Index (*H'*) across sampling regime at different farms. Arable (A), beef (B), dairy (D) and mixed grazing (MG). Shaded area displays the kernel probability density. Large black dot indicates median *H'* for each group.

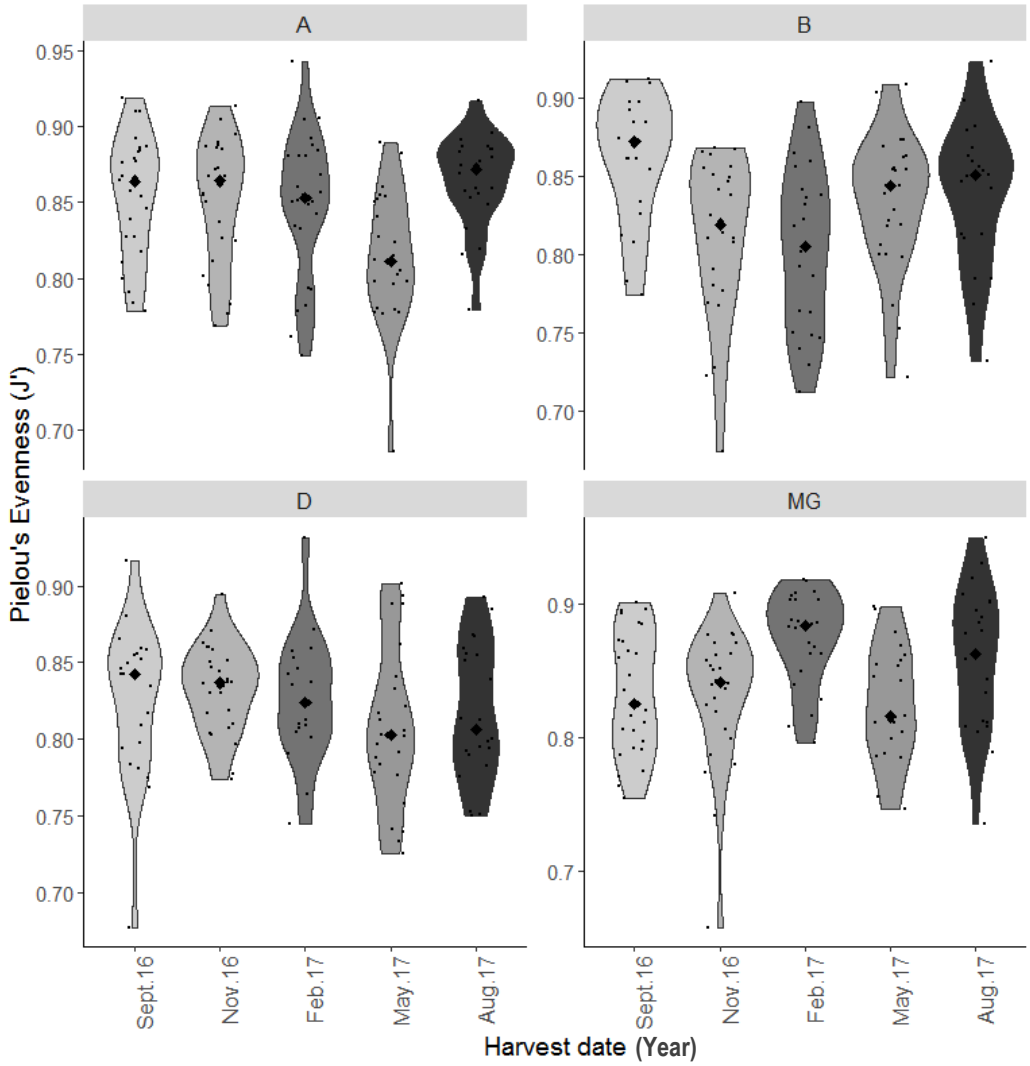


Figure 6.5 Trimmed violin plots showing Pielou's Evenness (J') across sampling regime at different farms. Arable (A), beef (B), dairy (D) and mixed grazing (MG). Shaded area displays the kernel probability density. Large black dot indicates median J' for each group.

6.4.3 Peak Identification

Peaks were identified by cross-reference to the peak library generated *in silico* (Supplementary Table 3) and to peak profiles generated for positive pathogen controls, compensating for calculated size drift (see Chapter 5) and allowing for a ±1 bp error margin. When a peak size matched to the predicted dominant peak for a pathogen in the database, attempts were made to locate other sub-peaks in the sample profiles which could further support the identification. Matches were found between peaks and pathogen dominant peaks, indicating the potential presence of numerous pathogens, including; *C. baratii, C. botulinum, C. chauvoei, C. perfringens, C. sporogenes, C. novyi* and *Paeniclostridium sordellii.* Sub-peaks could be tentatively identified for *C. novyi, C. sporogenes* and *Pn. sordellii,* but not for other species. A summary of the prevalence (%) and average RA, for suspected pathogen-peaks is shown in Table 6.4.

6.4.4 Correlation between peaks

Spearman's rank correlation matrixes were generated for main 6FAM, VIC and NED peaks, respectively. Significant, positive correlations were found between 6FAM (Supplementary Figure 7), VIC (data not shown), and NED peak groups (Supplementary Figure 8). Some of these correlated peak groups were selected for modelling analysis to investigate the response of peak abundance to measured variables.

Table 6.4 Summary of prevalence, abundance and identification of putative pathogens peaks

Peak	Dye	Prevalence (%)	RA[1]	Peak match	Expected peaks
546	6FAM	87.34	1.80	*Pn. sordellii*	547
576	6FAM	84	1.58	*Pn. sordellii*	577
595	6FAM	65.26	1.17	*Pn. sordellii*	596
340	VIC	32	0.79	*C. botulinum* Group I	341
346	VIC	65.89	2.20	*C. sporogenes/botulinum* Group I	345
416	VIC	16.21	1.05	*C. botulinum* Group I	417
459	VIC	12	0.67	*C. perfringens*	459-460
491	VIC	80.63	2.37	*C. botulinum*	491-493
496	VIC	57.47	0.81	*C. botulinum* Group I	495-496
504	VIC	11.37	1.05	*C. chauvoei*	503
529	VIC	33.26	0.87	*C. baratii*	529
344	NED	65.26	4.08	*C. novyi* Type A	344
403	NED	75.79	4.02	*C. novyi* Type A	402
424	NED	86.11	6.54	*C. novyi* Type A/B	423

[1] Relative abundance calculated as the average relative peak abundance for non-zero sites

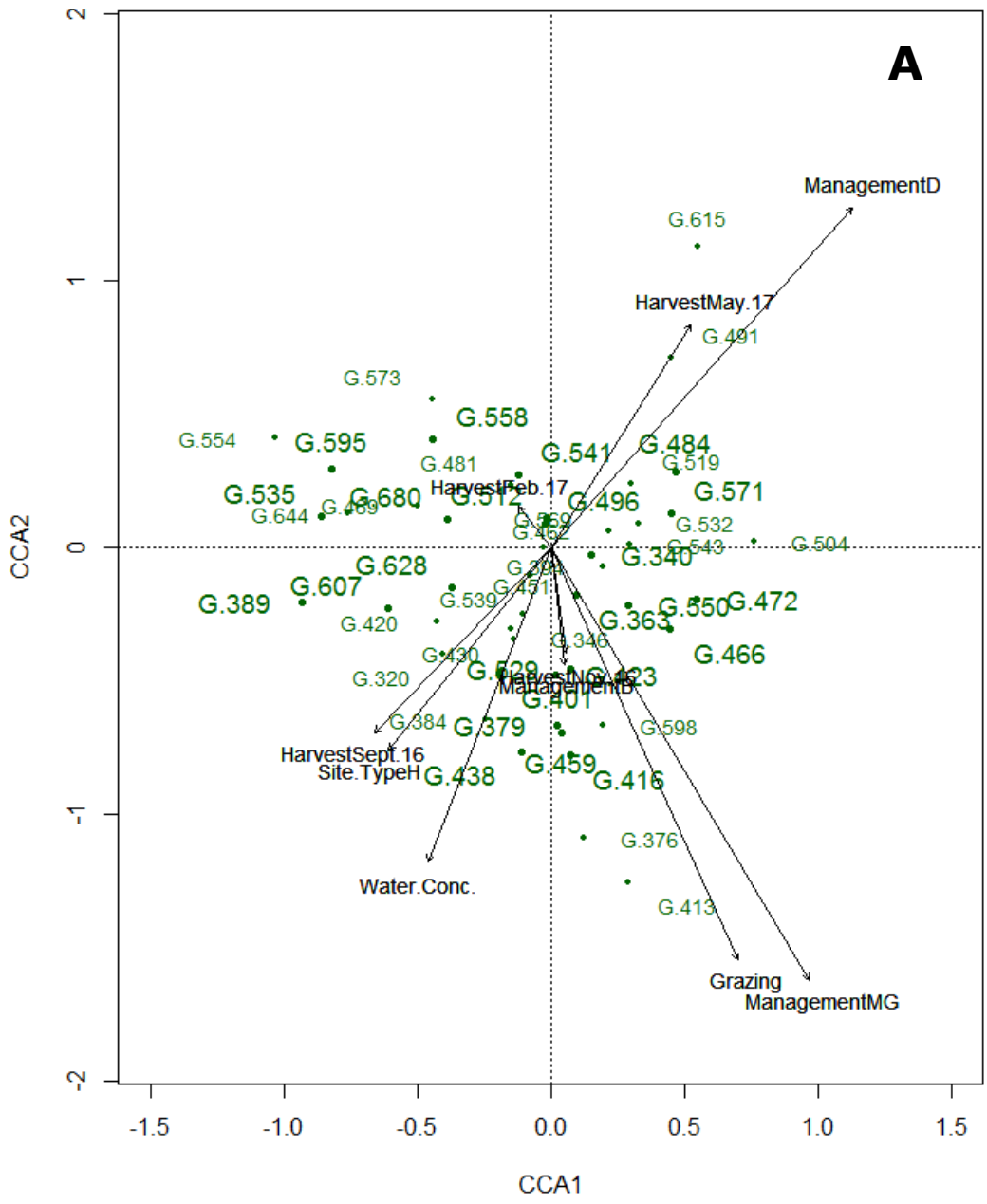
6.4.5 Canonical correspondence analysis

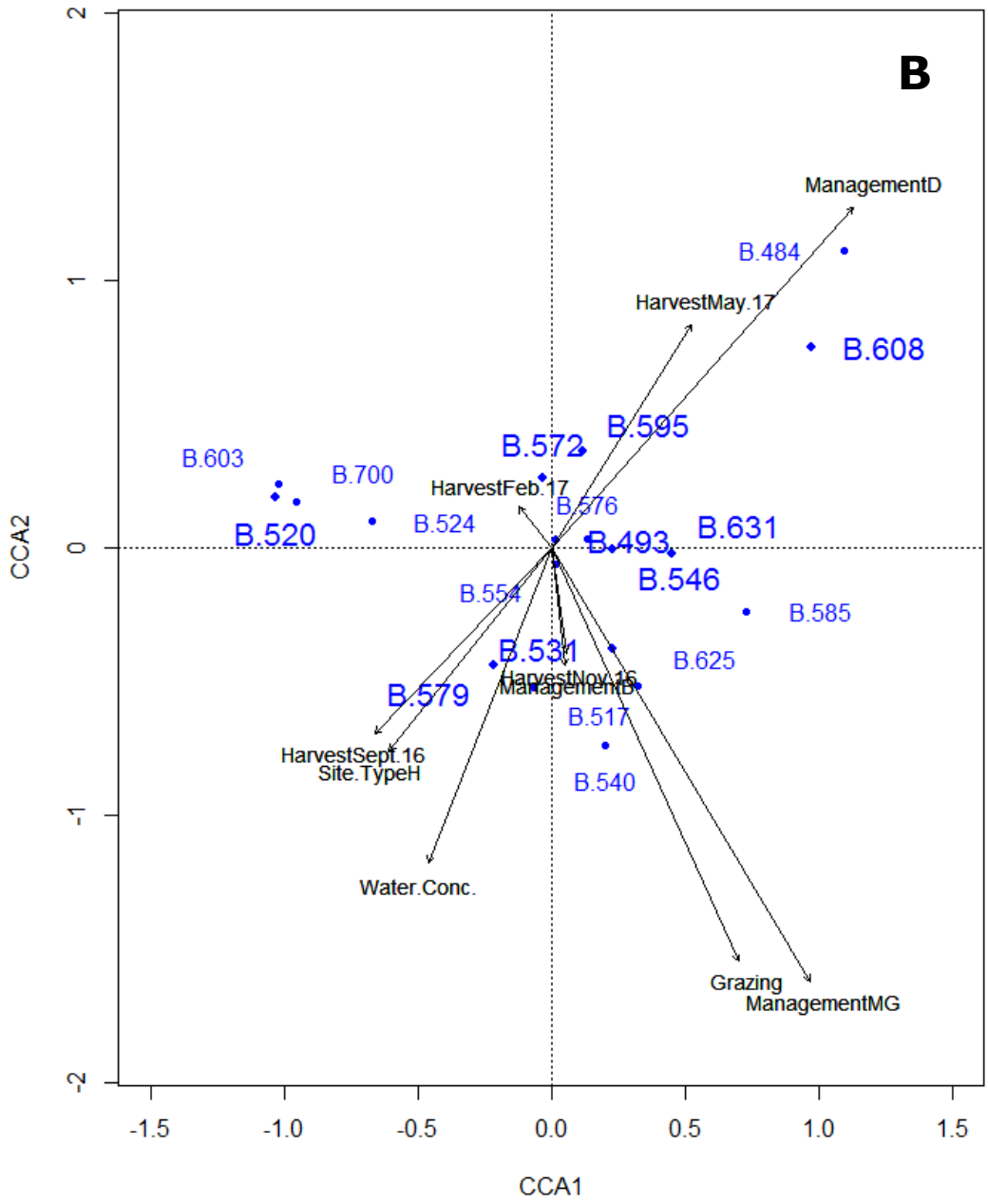
CCA and partial CCA (PCCA) was performed on RA data for 84 main peaks (present in over 10% of sites). This constituted 19 6FAM-labelled, 48 VIC labelled, 16 NED-labelled and 1 PET-labelled peaks. Figure 6.6 shows the CCA plots for each separate dye trace. The plot for the PET trace is not shown due to the lack of peaks. In the full CCA model, predictors explained 18.4% of the total variance of the data. The first constrained axis (CCA1) accounted for 16.8% of the constrained variance, and the second (CCA2)

12.6 %. Permutation tests indicated that the model (f = 6.30, p = <0.001*),* all predictors (p = <0.001), and both CCA1 (f = 17.99, p = <0.001*)* and CCA2 (f = 8.38, p = <0.001*)* axis were significant (Table 6.5). PCCA was then performed to assess the effect of the environmental variables with the inter-farm dependent variation removed. The PCCA model only accounted for 10.7 % of the total variance in the model, emphasising the considerable effect farm management and other inter-farm differences have on the clostridial community (Table 6.5). Ordination plots for the VIC trace showed dispersed peak scores across the plot, with little discernible clustering between peaks (Figure 6.6A). The 6FAM trace plot showed the close clustering of some peaks (Figure 6.6B). CCA indicated that some of the measured environmental variables and management factors influenced the bacterial community detected by this trace; putative *Peptostreptococcaceae* spp. Similarly, some peaks on the NED trace correlated with the management variables (Figure 6.6 plot C).

Table 6.5 Summary table for ANOVA-like permutation tests for CCA and pCCA

	CCA			PCCA		
Term	**Chi²**	**F statistic**	**p-value**	**Chi²**	**F statistic**	**p-value**
Model	0.522	10.47	<0.001	0.220	6.30	<0.001
CCA1	0.168	33.77	<0.001	0.090	17.99	<0.001
CCA2	0.126	25.21	<0.001	0.042	8.38	<0.001
Management	0.302	20.18	<0.001			
Harvest date	0.108	5.39	<0.001	0.108	5.39	<0.001
Soil water	0.038	7.53	<0.001	0.038	7.53	<0.001
Grazing	0.034	6.72	<0.001	0.034	6.72	<0.001
Site type	0.041	8.29	<0.001	0.041	8.29	<0.001
Residual	2.313			2.313		


A
ManagementD
G.615
HarvestMay.17
G.491
G.573
G.558
G.554
G.595
G.541
G.484
G.481
G.519
HarvestFeb.17
G.535
G.680
G.512
G.496
G.571
G.644
G.469
G.532
G.504
G.628
G.340
G.607
G.389
G.539
G.550
G.472
G.420
G.363
G.346
G.466
G.320
G.401
G.384
G.379
G.598
HarvestSept.16
Site.TypeH
G.438
G.459
G.416
G.376
Water.Conc.
G.413
Grazing
ManagementMG
CCA1
CCA2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2
1
0
-1
-2


B
ManagementD
B.484
HarvestMay.17
B.608
B.595
B.572
B.603
B.700
HarvestFeb.17
B.576
B.520
B.524
B.493
B.631
B.546
B.554
B.585
B.531
B.625
HarvestNov.16
ManagementB
B.579
B.517
HarvestSept.16
Site.TypeH
B.540
Water.Conc.
Grazing
ManagementMG
CCA2
CCA1
2
1
0
-1
-2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

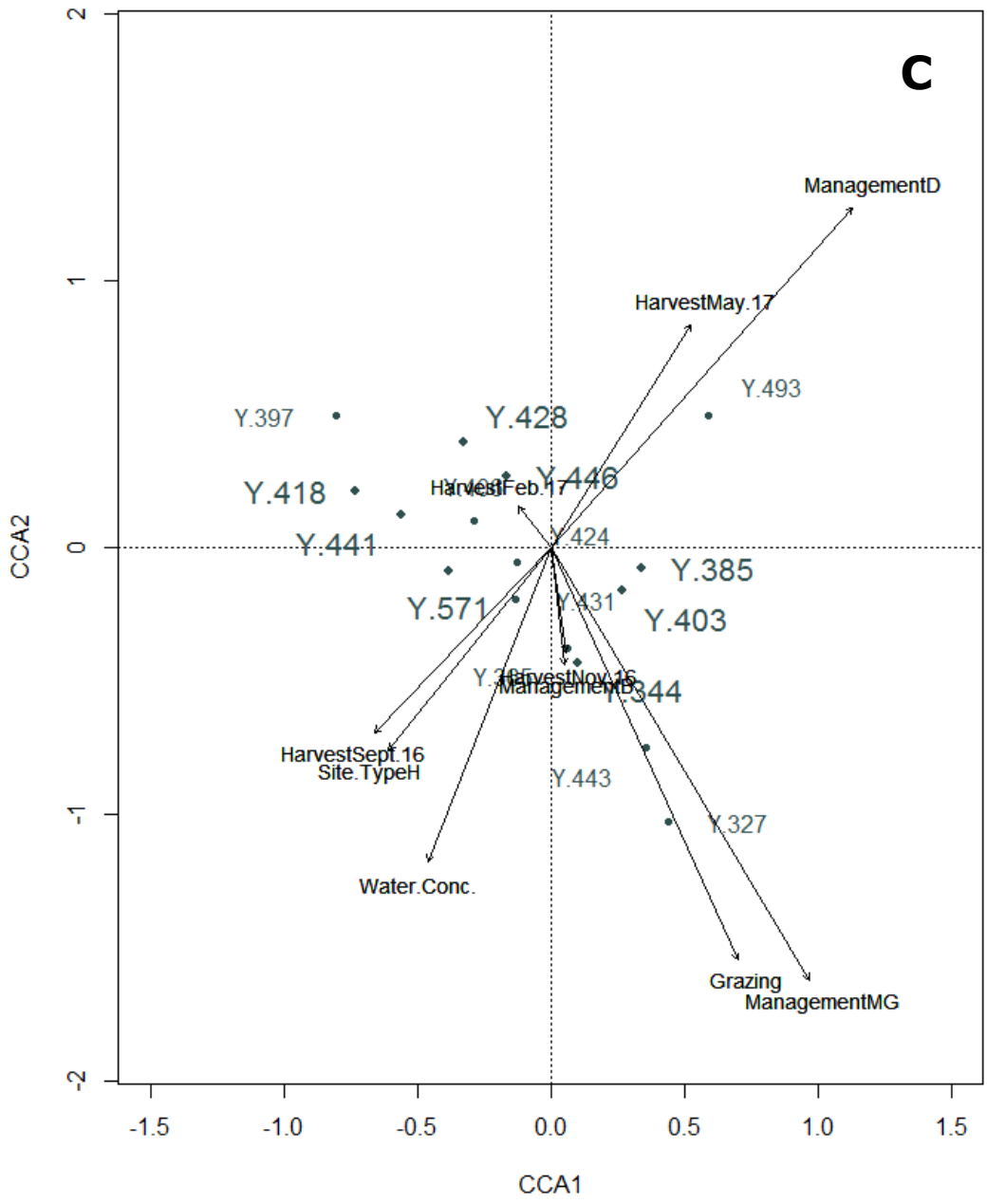


Figure 6. Error! No text of specified style in document.**6 Canonical correspondence analysis (CCA) plots for main peaks (found in >10% of samples).** The black lines and labels denote management factors, sampling periods and soil variables. Peak scores indicated by points, with peak labels (dye and length in bp) offset for clarity. Plots show ordination for VIC (**A**), 6-FAM (**B**) and NED (**C**) labelled peaks respectively.

6.4.6 Response of peak relative abundance to predictors

Using negative binomial hurdle models, the presence/absence and RA of suspected pathogen peaks and other selected peaks was found to be influenced by the measured variables. Negative binomial hurdle models consistently gave a better representation of the RA data for suspected pathogen peaks (in terms of lower AIC scores and correct zero-prediction) when tested on a subset of peaks. For some peaks which were present in all or most samples, negative binomial models were used.

6.4.7 6FAM peaks

Peaks could be categorised into 3 peak clusters, likely representing distinct *Peptostreptococcaceae* species, based on similar presence and changes in RA in response to different predictors, and positive, significant RA correlation between peaks (Supplementary Figure 7). Summary parameter estimates for predictors and relevant interactions, for selected peaks are given for the positive abundance (Supplementary Table 4A) and for the presence/absence (zero-count) model part (Supplementary Table 4B). For the positive abundance part of the models, RA of 6FAM Peaks 493, 546, 576 and 595 (putative *Pn. sordellii*) generally increased under beef and dairy management regimes, decreased with grazing intensity, and was lower in wet hotspots (Supplementary Table 4A). These effects were similar for the binary part of the model, meaning the probability of detecting these peaks at a site was significantly increased by livestock farming, but decreased in wet hotspots sites and with more grazing (Supplementary Table 4B). The change in RA for selected 6FAM peaks in wet/dry spot areas, under different management regimes is shown in Figure 6.7. A second group, including Peaks 520, 524, 579, 603 and 700, were only prominent in wet spots under arable cultivation. Significant increases in RA were associated with wet hotspots (Supplementary Table 4A). The third group contained Peaks 531 and 625.

6.4.8 VIC peaks

The were no similar trends between VIC-labelled peaks to indicate they belonged to the same species, and very few positive and significant correlations between RA were found (data not shown). All suspected *C. botulinum* Group I/ *C. sporogenes* peaks (except 340) demonstrated some degree of seasonality (Figure 6.8) and change in abundance between management types (Figure 6.9). Peaks 340 and 491, putatively associated with Group I *C. botulinum,* were more like to be detected and were more abundant at the dairy sites. Peaks 340 and 346 had a higher probability of detection in September and November, while Peaks 491 and 496 had a lower probability of detection from September through to February (Supplementary Table 5B). When significant, the RA and probability of detection was lower in wet spots for this group of peaks. For Peak 459 (putative *C. perfringens*), probability of detection fell over February, and was higher at beef and mixed grazing farms comparatively to arable (Supplementary Table 5B). The RA of Peak 504 (tentative *C. chauvoei*) was significantly increased in February and May, and the probability of the peak being present increased with grazing. When present, RA increase of Peak 529 (putative *C. baratii*) was associated with February harvest, wet spots, increased grazing and increased grazing on wet soils (Supplementary Table 5A; Figures 6.8 and 6.9).

6.4.9 NED peaks

Presence and abundance coefficient summaries for three peaks, putatively identified as *C. novyi,* are given in Supplementary Table 6. When present, Peaks 344, 403 and 424 showed some seasonality with higher abundance evident in November and under mixed grazing regimes (Figure 6.10). The probability of being present in a soil slightly but significantly increased with higher relative soil moisture content (Supplementary Table 6) but was lower at all other harvests compared to August 2018.

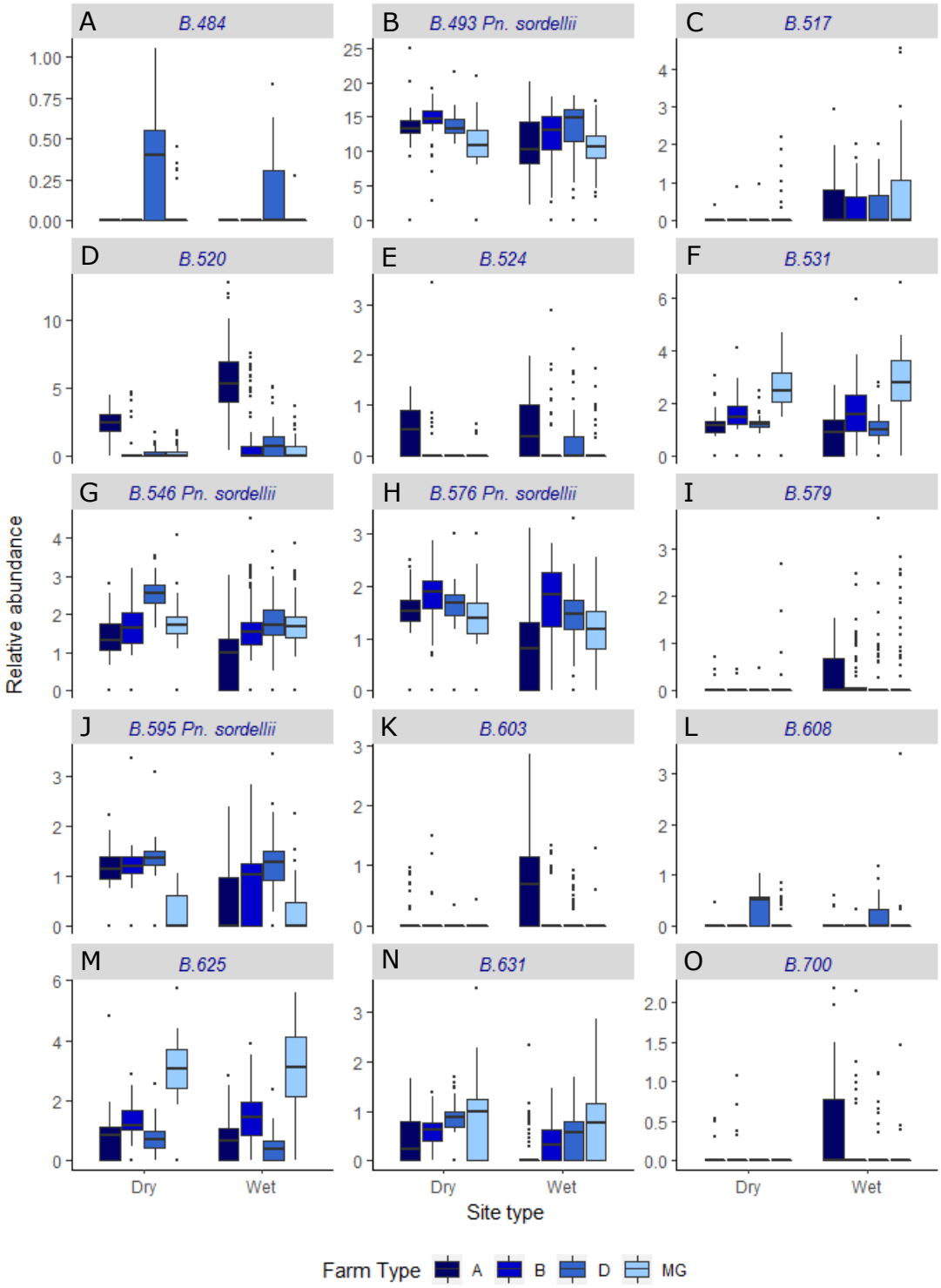


Figure 6.7 Boxplots showing the relative abundance of selected 6FAM peaks in wet and dry spots under different agricultural management practices. Farm types: Arable (A), beef (B), dairy (D) and mixed grazing (MG).

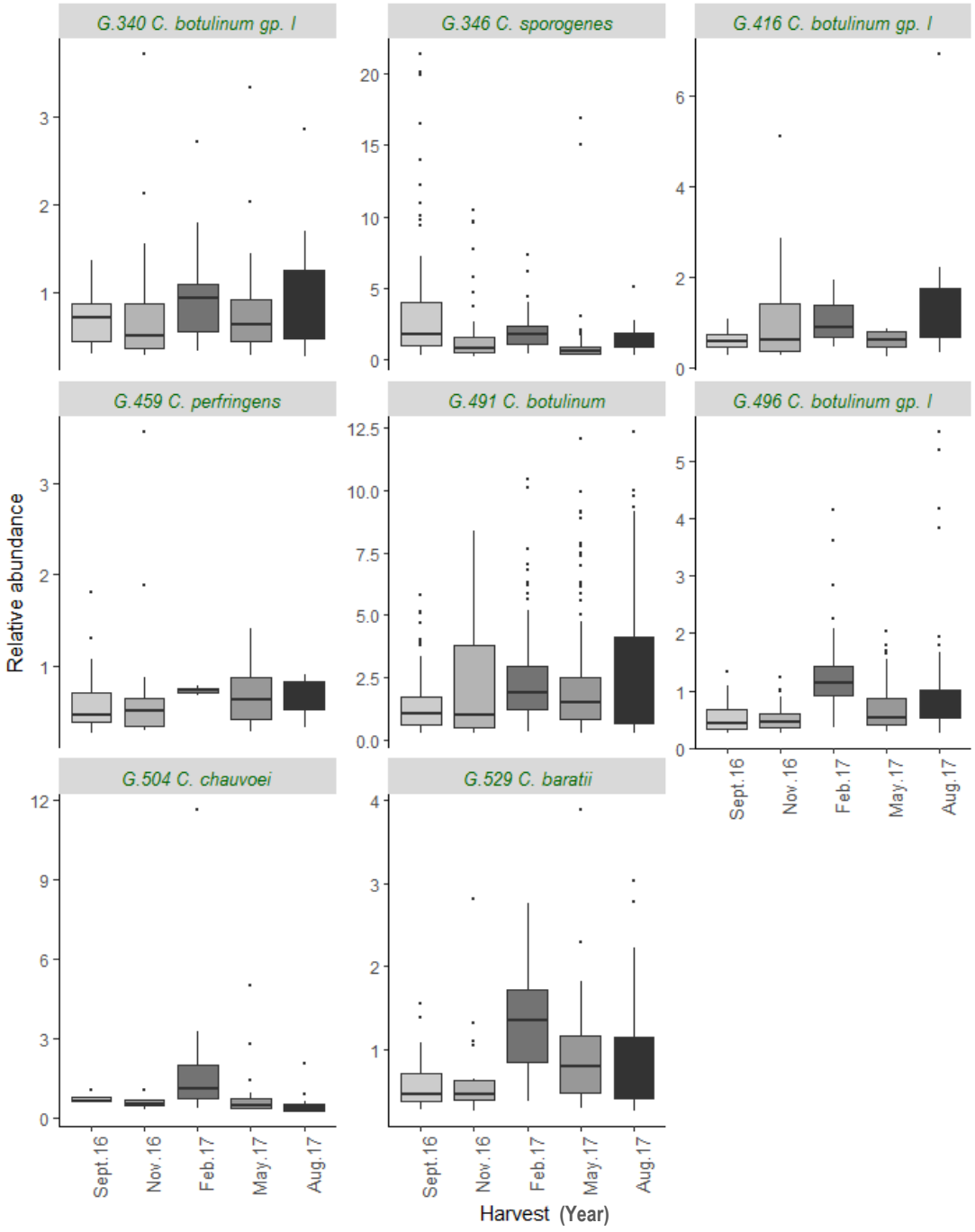


Figure 6.8 Boxplots showing the relative abundance of selected VIC-peaks across the sampling period. Zero counts (absent sites) were removed from the analysis. Farm types: Arable (A), beef (B), dairy (D) and mixed grazing (MG).

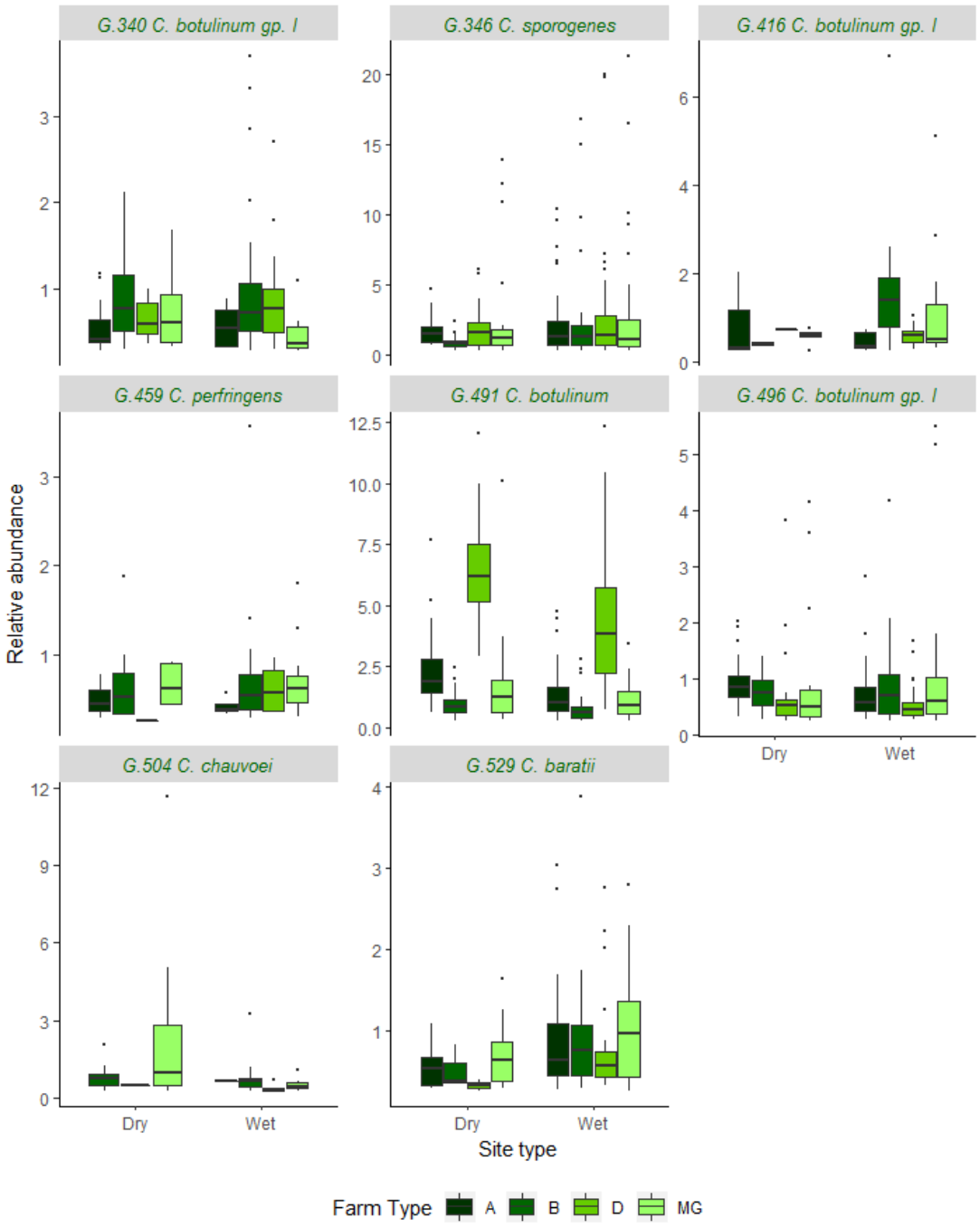


Figure 6.9 Boxplots showing the relative abundance of selected VIC-peaks in wet and dry spots under different agricultural management practices. Zero counts (absent sites) were removed from the analysis. Farm types: Arable (A), beef (B), dairy (D) and mixed grazing (MG).

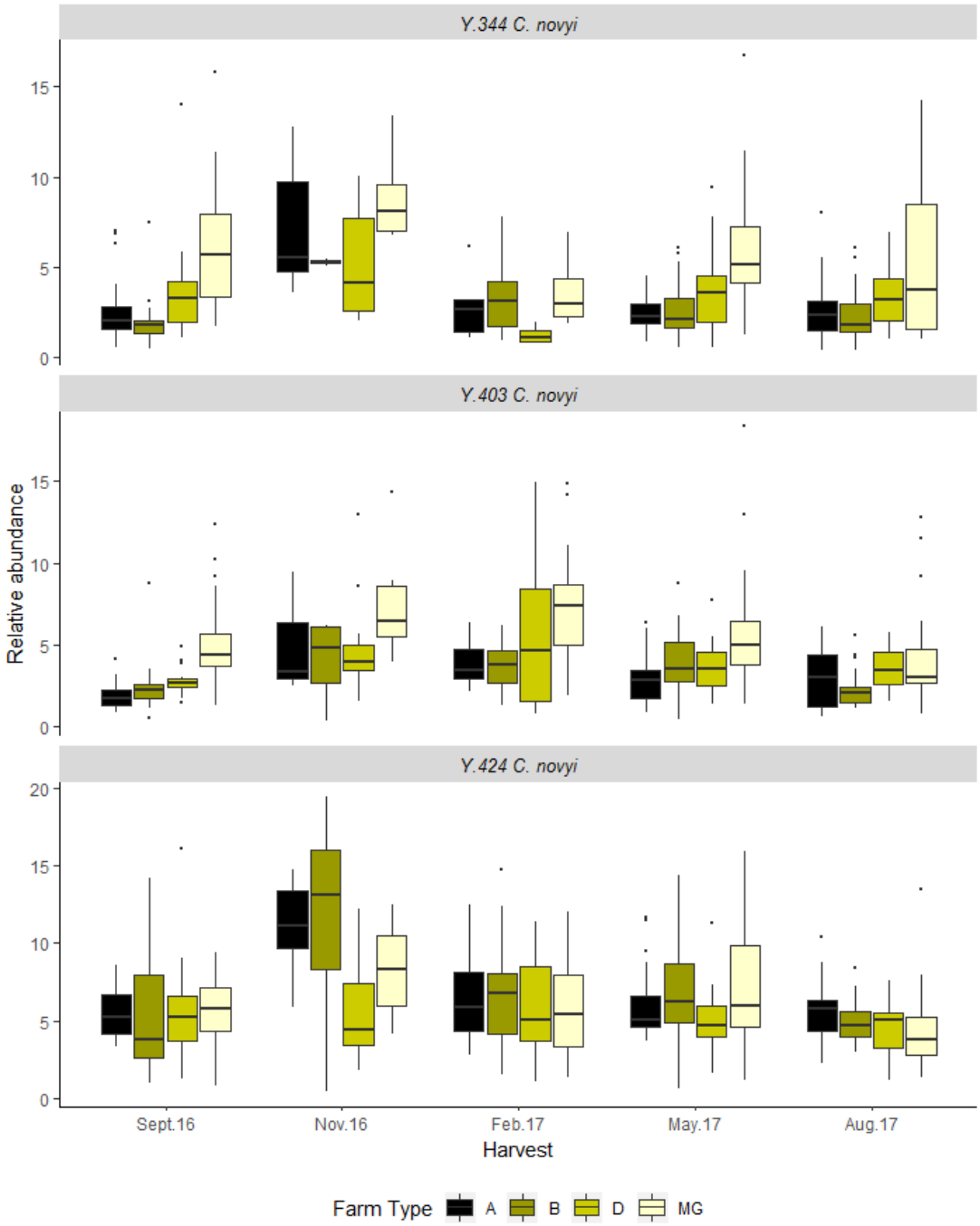


Figure 6.10 Boxplots showing the relative abundance of selected NED-peaks across the sampling period under different agricultural management practices. Zero counts (absent sites) were removed from the analysis. Farm types: Arable (A), beef (B), dairy (D) and mixed grazing (MG).

6.5 DISCUSSION

6.5.1 ARMISA method for clostridial analysis

This study trialled a novel molecular assay to investigate the presence and behaviour of clostridial pathogens in Scottish soils. The cost of multi-pathogen screening with molecular techniques at this scale is normally prohibitive, so to overcome this, an ARMISA protocol utilising PCR to amplify the 16S – 23S rDNA intergenic spacer region (ITS) was used. This region allows greater distinction between species and strains than is possible using more conserved genetic targets, such as the 16S rRNA gene. The protocol uses multiplexing to only target a small group of clostridial species which can then be rapidly identified using automated amplicon size calling and visualisation, as discussed in Chapter 4. The automated step also determines the relative fluorescence of each fragment, allowing the relative abundance of target organisms to be ascertained. This gave this study an advantage over previous presence/absence studies (Vidal *et al*., 2013; del Mar Gamboa, Rodríguez and Vargas, 2005; Lúquez *et al*., 2005), as the relative abundance of suspected pathogen peaks in soil samples could be calculated, and the change in relative abundance in response to environmental and management variables modelled.

One practical concern was of the likelihood of false negative results due to low DNA extraction efficiency from bacterial spores, and low-level Clostridia presence being below the limit of detection of the assay. An additional freeze-thaw step was added to the Mobio Powersoil extraction protocol, as recommended by the manufacturer, to help increase spore lysis and increase DNA yield. A modification to the AMRISA protocol detailed in Chapter 4 was applied by splitting the mPCR with 6 primers into 2 smaller mPCRs, with 3 and 4 primers in each, respectively. The purpose of this was to increase the sensitivity of the method by circumventing the issues associated with large multiplex reactions (Parker *et al*., 2015). Indeed, suspected Clostridia were detected in 533 (98.7%) of the soils screened.

Good quality profiles were generated for 475 samples (88.0% of total soils extracted) which were further analysed for pathogen identification and behaviour, and community analysis. Multiple, clear peaks were categorised after removing peak 'noise' and non-informative peaks. The PET channel only yielded 800 informative peaks (0.84% of total detected peaks), whereas the 6FAM channel generated 5399 good-quality peaks (95.1%). The increased noise with PET may be due to spectral bleed-through between dye channels, which has been observed in other studies using the same proprietary dye set (Hill, Butler and Vallone, 2009). No-PCR clean-up stage was used in this methodology. In Chapter 5, a simple ethanol-precipitation clean-up step gave no observable improvement in peak-profile quality. This step was therefore removed here to improve the speed and cost-effectiveness of the protocol. However, the use of alternate amplicon-cleaning techniques could resolve the excess noise sizing quality issues observed.

To attribute peaks to species groups and model their respective relative abundance, peaks that were present in less than 10% of site profiles were removed. This reduced the number of peaks to 84 main peaks. After the removal of rare species, only one PET-labelled peak remained in the dataset. The PET trace was designed to only target *C. tetani* spp., and the predicted *C. tetani* peak-sizes did not match the one peak observed. The sensitivity and specificity of this channel was not tested *in silico* (see Chapter 5), and it is possible that PET-labelled primer was not specific to the target group. A better solution may be to forfeit the detection of *C. tetani* and use this dye channel for the detection of an Internal Amplification Control (IAC). By running positive controls simultaneously to sample mPCRs, it was possible to rule out PCR failure due to thermal cycler malfunction, incorrect PCR mixture and poor polymerase activity, however an IAC could aid in the detection of inhibitory compounds.

6.5.2 Peak identification

After accounting for drift and allowing small (±1bp) discrepancies between observed vs predicted peaks, 14 peaks were tentatively matched to 6 pathogenic species, including *C. baratii,* Group I *C. botulinum/C. sporogenes, C. chauvoei, C. perfringens, C. novyi* and *Pn. sordellii.* The identification of *C. novyi* and *Pn. sordellii* was further supported by the presence of additional sub-peaks in the same samples. These sub-peaks clustered together in CCA ordination plots and showed similar distribution patterns across sites, indicating that they likely originate from the same organism. Based on the 6FAM-labelled 546 bp peak, *Pn. sordellii* was present in 87.34 %, and on NED-labelled 424 bp peak, *C. novyi* in 86.11% of profiles, respectively. This is in good agreement with del Mar Gamboa *et. al* (2005) (del Mar Gamboa, Rodríguez and Vargas, 2005), who identified both pathogens as prevalent pathogens, albeit in Costa-Rican soils that were non-agricultural. Caution was exercised regarding the inference of pathogen presence from only one matching peak. There are numerous clostridial species which have not been fully genome sequenced, and hence are not included within the peak database. Such species could theoretically produce identical peak sizes to pathogen species, resulting in false-positives peaks. Confirmation with 2 or more peaks will substantially reduce the chance of false-positive peak identification. Likewise, identification of many of the peaks is currently hindered by the lack of publicly-available fully-assembled genomic data. Many species, including the pathogens *C. haemolyticum, C. intestinale* and *C. septicum,* are underrepresented in terms of available genomic material, making accurate *in silico* peak prediction problematic. As more material is released, the AMRISA protocol will be more effective at identifying Clostridial species.

6.5.3 Inter-farm and temporal variation of clostridial community

Changes in the clostridial community were assessed using the total peak number (*S;* before removal of rare peaks), Shannon index (*H'*) and Pielou's evenness (*J'*). It is stressed that these metrics should not be used to infer

true diversity, or even clostridial diversity. As discussed, the design of the AMRISA test is such that each true species can give rise to multiple peaks, which would lead to an overestimation of the diversity by these metrics. Furthermore, the test is designed to target pathogenic species and select-closely related species, meaning only a fraction of the true clostridial population is detected. The routine processing required for fragment analysis data, such as noise removal, can remove rare species leading to further deviations between the measured and true diversity (Blackwood *et al*., 2007). These metrics were used to indicated changes in the pathogen and pathogen-like clostridial community in response to environmental and management factors.

The total number of unique peaks detected, and the Shannon Index were significantly lower in February than other months. Other studies have also documented a decrease in generic bacterial diversity in agricultural soils over winter months (Smit *et al*., 2001). This was the coldest part of year with daily average temperatures regularly below 5°C for the 2 months prior to sampling. The fall in species richness in this current study indicates that cold resilience is not ubiquitous across all clostridial representatives. Indeed, Brochier *et al.* (2012) observed a decline in *C. perfringens* in French soils during a similar cold period.

Although total peak number was lower in dairy and mixed grazing sites than in arable sites, increased grazing intensity significantly increased the total number of peaks. This suggests that as grazing levels or stocking density increase, the more faecal matter is added to the soil and more clostridial spp. are spread onto the soil. Yang *et al.* (2013) found that the abundance of Firmicutes *amoA* and clostridial virulence, stress and antibiotic resistance genes also increased in response to grazing, respectively. The changes in clostridial communities between farms observed here may be due to soil physicochemical differences between farms. However, total peak number and evenness did significantly increase during February and May under some livestock regimes.

6.5.4 Change in pathogen presence and abundance

The findings here show how the relative abundance of some putative pathogens differed depending on season, farm management practices, prolonged soil saturation and grazing intensity. The pathogen *Pn. sordellii* had a higher relative abundance in soils collected from dairy and beef sites. This pathogen has been isolated from cattle manure in other studies (Bagge, Persson and Johansson, 2010), and is likely introduced to the soil by cattle manure. *Pn. sordellii* has been increasingly reported as the cause of severe abomasitis in calves and lambs in Scotland, and changes in land-management might be required to reduce disease risk from highly-contaminated soils. The high prevalence and abundance of the pathogen across all sites, including at arable farms, and in non-agricultural soils. This indicates that the pathogens readily naturalise, are particularly resistant to destruction (in spore form) or are a prevalent in the common soil microflora. Relative abundance decreased in wet hotspots, which may be due to increased competition from more specialised non-pathogenic clostridia not detected by the test applied, or increased predation of vegetative cells, as discussed in Chapter 4. The behaviour of two other putative *Peptostreptococcaceae* spp. was also described. Although they could not be attributed to a pathogenic species (which, in part, is due to the very few genomic sequences available for this family), they demonstrate clear presence and abundance responses to environmental variables, further validating the utility of AMRISA as a tool for clostridial surveillance.

C. novyi appeared to show strong seasonality, with relative abundance higher in November and February. *C. novyi* is the causative agent for highly fatal black disease (infectious necrotic hepatitis) in sheep and cattle. Black disease normally follows damage to hepatic tissue by liver fluke; parasitic trematoda (*Fasciola hepatica*) which are associated with wet, boggy soils. Liver fluke has traditionally been prevalent in the west coast of Scotland, due to the wetter soil conditions. However, warmer and wetter winters are

thought to be responsible for the recent spread of *F. hepatica* into previously fluke-free farms in southern and eastern Scotland (Kenyon *et al*., 2009). Black disease is thought to be caused by ingestion of clostridial spores from pasture, which pass into the liver where they remain dormant until tissue damage occurs. Therefore, high risk periods for *C. novyi* ingestion may not correspond to the onset of disease. The observations in this study show that suspected *C. novyi* abundance was greater in mixed-grazing systems during wetter months of the years, which could lead to an increased concurrent risk of both liver fluke and black disease in these areas, which may have been previously low-risk. Similarly, Peak VIC-529, tentatively identified as *C. baratii,* showed increased relative abundance and probability of detection in wet hotspots and intensively grazed sites. Peak VIC-504, putatively identified as *C. chauvoei* was virtually absent from sites under arable management, indicating that the bacteria are probably excreted in faeces and accumulate in the soil. This is in agreement with other research which suggests that pastures can become heavily contaminated, leading to yearly reoccurrence of Blackleg (Bagge, Lewerin and Johansson, 2009; Hang'ombe *et al*., 2000).

6.6 CONCLUSIONS AND FUTURE ASPECTS

This study has demonstrated that clostridial species including pathogenic representatives, are prevalent in farm soils in the East of Scotland. Pathogenic species are not necessarily ubiquitous in the soil, but rather their prevalence and abundance are affected by management regime, specifically the type and intensity livestock grazing. Seasonality influenced the abundance and composition of clostridial communities, and some significant interactions between season and farming type on pathogen abundance were identified. *Paeniclostridium sordellii* was one of the most prevalent and abundant clostridial pathogens identified, with higher abundance observed in cattle-grazed soils. *Clostridium novyi* was also highly prevalent and abundant across sites, with higher abundance noted

in winter sampling months. Other important pathogenic species tentatively identified were *C. baratii, C. botulinum* (proteolytic) and *C. chauvoei.* The multivariate study allowed for a better understanding of temporal behaviour of clostridial species and communities than can be derived from presence-absence only studies. This raises questions about the impact of intensive livestock systems on the microbiological quality of the soil, and the change to disease risk this poses to humans and livestock. Additionally, climatic influence on pathogen behaviour was demonstrated, which needs to be considered in light of global warming and the potential effect this may have on disease epidemiology. As more genomic sequences become publicly available, it is anticipated that many of the currently unidentifiable peaks detected can be identified. Future epidemiological studies can utilise this tool to screen other environmental reservoirs where clostridial pathogens may be present, such as marine and estuarine sediments. Future work should seek to complement multi-pathogen surveillance with specific, targeted molecular tools such as quantitative PCR, allowing a more sensitive and accurate pathogen enumeration than was possible in this study, while further focus on the effect soil type, structure and chemistry on clostridial behaviour is still needed.

CHAPTER 7: GENERAL DISCUSSION

7.1 METHODS FOR THE DETECTION, IDENTIFICATION AND ENUMERATION OF CLOSTRIDIAL PATHOGENS

The literature review highlighted two issues which needed addressing prior to the testing of growth-related clostridial hypotheses in this work, and indeed for future studies. Firstly, there was a clear need for standardisation and harmonisation of diagnostic techniques for clostridial detection (Fang, Polage and Wilcox, 2017). Secondly, it was apparent that the existing taxonomic definitions for *Clostridium* and Clostridia species were antiquated, with many nomenclatural discrepancies between databases. This is demonstrated by the multi-antibiotic resistant pathogen *Clostridioides difficile,* which was renamed from *Clostridium difficile* to *Peptoclostridium difficile* (Yutin and Galperin, 2013)*,* then to *Clostridioides difficile* (Lawson *et al*., 2016)*,* all within the duration of this project. This hindered the procurement of sequence and literature resources, as many could not be identified or retrieved using one species name as a search term. Additionally, recent genomic studies indicate that the nomenclature of some clinically-important strains does not reflect their true phylogenetic relationships (Skarin and Segerman, 2014; Weigand *et al*., 2015). An important example is the distinction between Group I *C. botulinum* and *C. sporogenes.* The latter had always been considered a closely-related relative of *C. botulinum* (Group I)*,* that was distinguished by the lack of toxicity (and absence of toxin genes). Indeed, *C. sporogenes* is commonly used as a safe surrogate for the former in food and clinical studies. A study published in 2015 showed that while both species cluster into distinct phylogenetic clades, the neurotoxic *bont/B2* gene can be gained and lost by both species through horizontal gene transfer (Weigand *et al*., 2015). This phenomenon has inadvertently led to the incorrect identification of many strains of both species, which could have serious ramifications for human health. Similarly, Skarin and Sergerman (2014) used genome and

plasmidome analysis of the animal pathogens *C. botulinum* (Group III)*, C. novyi and C. haemolyticum* to reveal many misclassified strains and demonstrated phage-driven transfer of plasmids and toxin genes between these closely related species. These results had implications for this research as it was possible that some of the genomic data that had been retrieved were misclassified, and furthermore, it casts serious doubt on whether toxin-based detection and identification methods are suitable for Clostridia.

To ascertain the true evolutionary relationships of fully-sequenced pathogenic clostridial species, a multilocus approach was devised, as described in Chapter 3. Homologous nucleotide sequence for 50 conserved housekeeping genes were retrieved, concatenated, and phylogenies were inferred using both maximum likelihood and Bayesian inference models. The phylogenies indicate the misclassification of *C. aceticum,* which fell outside the *Clostridium sensu stricto* clade, and the misnaming of *C. botulinum* Prevot 594 and *C. pasteurianum* NRRL B-598. This outcome supports the body of literature encouraging the use of phylogenetic comparison as the universal method of taxonomic assignment. This methodology was used as an alternative to whole-genome comparisons techniques, such as the genome-wide Average Nucleotide Identity (gANI) metric. Whole-genome comparison techniques have recently been recommended as the primary guide for new species assignment and for microbial species delineation (Varghese *et al*., 2015). However, software facilitating fast, multiple-genome comparisons have only come to the fore in recent years and can generally only utilise fully-assembled genome sequences. The methodology detailed in Chapter 3 has the advantage that it can utilise partially assembled genomes as contigs and scaffolds, although only fully-assembled sequences were used in this work. Furthermore, the reduced sequence length of the multilocus concatenation *vs*. the full genome allowed full phylogenetic reconstruction of 101 strains in under 5 hours. One alteration to the method described in Chapter 3 would be to retrieve translated nucleotide coding sequences using a protein

BLAST with the TBLASTN program (Altschul *et al*., 1997), then align sequences based on their amino acid sequence. This modification could improve the sensitivity of sequence retrieval and the accuracy of the alignment, which may result in a better phylogenetic inference for more distant branches in the phylogenetic trees (Figures 3.3A & 3.3B, (Abascal, Zardoya and Telford, 2010)). The *inf*B gene was identified as a suitable gene-marker to reconstruct the phylogeny of 238 *Clostridium* strains with partially-completed genome sequences (Figure 3.6). The clustering of clinical strains seen in this phylogeny emulated those inferred by whole-genome analysis used in the previously mentioned studies (see Figure 3.6;7, 8), validating the usefulness of this gene for accurate phylogenetic inference. Other genes were also putatively identified that could be useful genomic markers for future studies. It is proposed that future studies use such markers for the identification and taxonomic classification of new isolates prior to whole-genome sequencing. Sequencing this gene alone could indicate novel species and strains and is a more suitable genetic marker than the 16S rDNA for clostridial analysis.

For the same reasons, it was conceived that housekeeping genes could be targeted by diagnostic tools for the simultaneous detection and discrimination of clostridial pathogens. Other authors have successfully targeted these genes for similar purposes (Dhalluin *et al*., 2003), although assays reported in these studies could only discriminate to species level. As discussed in Chapters 2, 3 and 5, the clinical manifestation of clostridial disease is often dependent on the strain/pathotype of the pathogen, therefore diagnostic tools which allow strain distinction will have greater use to clinicians and epidemiologists. Whole-genome sequencing (WGS) is the ultimate tool for bacterial subtyping (Salipante *et al*., 2015; Saltykova *et al*., 2018), and in the near future it is probable that WGS will be best-practice for pathogen surveillance. However, at the current time, cost and technical skill requirements mean the technology is not universally applicable, particularly in developing nations where clostridial outbreaks are more prevalent. Nucleic acid amplification tests (NAATs) can be

performed in most clinical and research laboratories, and a well-designed PCR-based assay can be used to complement, or as an alternative to WGS. In this study, PCR-screening was favoured due to the study size, the co-contaminated nature of the samples, and the difficulties associated with isolating some clostridial pathogens using culture-based methods (Edwards, Suárez and McBride, 2013; Halm *et al*., 2010). The AMRISA protocol developed facilitated the simple detection and discrimination of 7 clostridial species tested *in vitro*, and most other clostridial pathogens that were tested *in silico.* Hitherto, a multi-species NAAT allowing simultaneous detection of all key pathogens and discrimination of different pathotypes has not been possible. To explore the utility of this test, agricultural soils and two anaerobic digestates were screened. Both soils and digestates were positive for various clostridial pathogens, including *C. perfringens* and *Cl. difficile.* In the UK, the industry standard for digestate microbial quality (BSI PAS 110) only specifies limits for *Escherichia coli* and *Salmonella* spp. in non-animal by-product digestates. Recent studies have also shown digestates can present a biohazard risk of pathogenic Clostridia (Neuhaus, Shehata and Krüger, 2015; Neuhaus *et al*., 2015), therefore upper limits should be included within the PAS 110 to reflect this. AMRISA could also be utilised for the routine screening of digestates, and other organic soil amendments. Future research should explore the application of this protocol to clinical and food DNA samples. In both settings, rapid, accurate pathogen identification is imperative, and the AMRISA tool could facilitate faster detection of contamination, aiding disease outbreak control. The proof of concept for AMRISA was demonstrated in Chapter 5, although further experimental work is needed to ascertain the sensitivity and specificity of the assay. The application of the protocol for environmental pathogen surveillance was demonstrated (Chapter 6), although its utility should improve further as more clostridial genomic sequences become publicly-available, allowing the identification of more peaks.

7.2 PREVALENCE AND BEHAVIOUR OF CLOSTRIDIA IN AGRICULTURAL SOIL

The overarching aim of this thesis was to elucidate the distribution, abundance and behaviour of pathogenic Clostridia in soils. In Chapters 1 and 2, the role of agricultural soils as a pathogen reservoir and vector of disease to food and animals was identified, particularly when organic and/or intensive grazing regimes are used. The focus of the research in this thesis was therefore designed to reflect this, with Chapters 4 and 6 exploring pathogen presence, persistence and diversity in farm soils. This was achieved using proprietary kits and in-house designed diagnostics, which facilitated the qualitative, and where necessary, quantitative measurement of clostridial pathogens. One current limitation of these molecular approaches is the inability to discriminate between DNA derived from cells in vegetative or spore forms, as discussed in Chapter 4. No suitable solutions to this problem could be found in the scientific literature, although transcriptomic or proteomic analysis of samples may provide a better understanding of pathogen physiological state (Trunet, Carlin and Coroller, 2017).

One key finding from this research is that not all clostridial pathogens are prevalent across soils types and locations. The spatial and temporal presence and abundance of Clostridia was investigated in Chapter 6, and variability was explained, in part, by differences in farm type, season and grazing intensity. This reinforces the view that farm management practices play a pivotal role in soil contamination and disease incidence. Future sampling efforts should consider screening a greater number of farms across more types of agriculture, although an immediate recommendation would be to encourage the use of crop and animal rotational practices. Higher abundances of specific pathogens were often associated with certain livestock regimes, therefore alternating between arable-livestock systems, or rotational grazing of different livestock might minimise the accumulation of certain pathogens and reduce the risk of reinfection in

high-risk areas. Clostridial communities remained consistent across replicates from each site, and therefore future sampling regimes may not need not to be as labour intensive. As a continuation from this project, it is recommended that live-culture techniques are used to isolate clostridial strains from the soils collected in this study. These strains could form the basis of an isolate collection of environmentally-derived Clostridia. DNA from these strains could be analysed individually using the AMRISA protocol and WGS approaches. This will aid further peak identification from the farm soils used in Chapter 6, permit the development of a reference peak-database for other researchers, and allow assessment of the complementarity and advantages of both approaches. Furthermore, it is predicted that many of the peaks identified in the farm soils could represent new clostridial species and strains, due to their dissimilarity to *in silico*-predicted peaks. New strains may represent new health-challenges, sources of antimicrobial resistance or potential uses within biotechnology.

When pathogens were spiked into soil microcosms under different water treatments, the driest soils permitted prolonged survival at elevated concentrations, whereas no detectable growth was seen in the saturated soils. The term 'saturated' was used to qualitatively infer the likely oxygen-deficient soil conditions, although measurement of soil redox potential (E_H) would have provided a more quantitative measure. Likewise, in the field soils, soil moisture content had little effect on the prevalence and abundance of the species (peaks) analysed. The use of *in situ* E_H measurement probes for field sites was not possible due to the concurrent farming practices, and it was felt that extracting soil pore water and measuring E_H was not a feasible approach given the other time requirements of the study. Interestingly, the effect of long-term waterlogging had a more pronounced effect on clostridial community dynamics and pathogen relative abundance. This suggests that some pathogens may have a competitive advantage in waterlogged soils, whereas others are outcompeted by indigenous, non-pathogenic species.

7.2.1 Implications for human health

In Chapter 4, real-time quantitative PCR was used for the sensitive enumeration of three pathogenic species co-spiked into soil. The spiking was intended to simulate clostridial contamination from faecal shedding or organic soil amendments, which are likely to be co-contaminated. The longevity of the three pathogens after inoculation (> 84 days) is important for hazard analysis and critical control points (HACCP)-based pathogen control. *C. perfringens* alone caused 177 deaths in England and Wales from food-borne illness between 1996-2000 (Adak, 2002). One of the risk reduction strategies for pathogen transfer from manured-soils onto crops is to apply manure at least 6 months prior to planting read-to-eat vegetables, and remove grazing animals 9 months before planting crops (Leifert *et al*., 2008). However, this research suggests that this time-period might not allow sufficient reduction of more resistant pathogens such as Clostridia and Mycobacteria (Ghodbane *et al*., 2014), particularly in dry-soils or drought conditions. The decay constants derived in Chapter 4 should be factored into quantitative-microbial risk assessments and HACCP assessments, particularly if unpasteurised or Clostridia-contaminated organic soil amendments are to be used.

One of the drivers for this study was to generate baseline data on the prevalence of clostridial pathogens in the soil. This data is key for risk assessment; land application of Clostridia-contaminated organic wastes to soils with pathogens already present, or to soils which can enhance the survival of introduced pathogens could increase the risk of disease exposure. The development of quantitative-microbial risk assessment models can be used to assess any increased risk to human health associated with increased or novel OSAs. However, the usefulness of these models is also dependent on the quality of baseline data for pathogens in soils and OSAs. This study showed the prevalence of human pathogen such as *Pn. sordellii* across farm soils and in anaerobic digestates. It is suggested that future work prioritises the quantitative assessment of the

clostridial pathogenic loads in OSAs and the subsequent pathogen load in the soil after short- and long-term application.

7.2.2 Implications for livestock health

Numerous putative animal pathogens were isolated from soils from different farm types. Simple presence/absence data for pathogens across individual farms could allow land-managers to make better informed decisions regarding grazing patterns. It was observed that the prevalence and relative abundance of many pathogens showed strong seasonality. Some peaks, such as *C. perfringens,* were less likely to be detected in February, whereas *C. novyi* peaks had a higher relative abundance over the wetter winter months, illustrating the influence of climate in soil-pathogen behaviour. This could have important implications for livestock disease patterns. Climate change observations for the UK predict a trend towards increasingly wetter and warmer winters (Ward and Hicks, 2013), and this in turn could result in changing patterns of clostridial disease. Diseases tended to be more strongly associated with regions subject to wetter climates and soil. An example of this is the high prevalence of braxy (*Clostridium septicum*) in Scotland (Lewis, 2011). This disease and black disease (also associated with wet, boggy soils) may become more widespread in the future. Indeed, parasitic sheep helminths, historically associated with wet soils in West Scotland, have recently been detected in soils in south-eastern Scotland (Kenyon *et al*., 2009). These parasites cause liver fluke, which results in damaged liver tissue susceptible to infection with *C. novyi* and *C. septicum* (Songer, 1996, 2010). As helminth prevalence increases in eastern Scottish regions, the wetter soils will also favour the growth of some clostridial pathogens, a synergy which it seems would inevitably lead to an increase in these clostridial diseases.

CHAPTER 8: FINAL CONCLUSIONS

In summary, this thesis has used a multidisciplinary approach to design, test and utilise phylogenetic and diagnostic tools to aid in the identification and detection of clostridial pathogens in agricultural soils. An updated clostridial phylogeny was produced, indicating important groupings of clinical strains and other taxonomic conflicts. Microcosm studies indicated the enhanced survival of *Cl. difficile, C. perfringens* and *C. sporogenes* in dry soils, contrary to what was predicted. However, a large-scale field study showed that inter-farm differences and sampling season had more pronounced effects on clostridial pathogen and community behaviour than soil water content. Grazing intensity and long-term waterlogging were also found to affect the prevalence and abundance of some suspected clostridial pathogens. The utility of the AMRISA protocol developed in this thesis for widescale surveillance of clostridial pathogens in environmental reservoirs was demonstrated and it could be utilised for other environmental, clinical or food focused research where clostridial contamination is expected.

Review existing literature on clostridial epidemiology and prevalence in soil to establish potential links between disease, pathogen behaviour and soil factors.

A comprehensive literature review was performed to identify the main soil, environmental and land management factors effecting the behaviour of clostridial pathogens in the soil (Chapter 2). Few studies to date have managed to statistically link higher pathogen prevalence to any underlying soil physicochemical parameters, although soil distribution patterns for *C. botulinum* have been established. Pathogenic Clostridia are, however, frequently isolated from sediments, marshes and rice-paddy fields, suggesting that increased soil water content causes soil anaerobicity, enhancing clostridial growth. This was supported by case-control and retrospective cohort studies which identified high rainfall, poor soil drainage and flooding events as risk factors for clostridial disease in animals, such as equine grass sickness and blackleg Increased grazing

intensity and meteorology-related factors were also associated with disease outbreaks. Furthermore, the type of land management and the use of organic soil amendments may affect pathogen abundance. In general, the existing pool of research is limited to small geographic regions, and the bulk of studies focus on just a few key species. The abundance and prevalence of *C. chauvoei, C. novyi, C. septicum*, *Pn. sordellii* and *Pr. bifermentans* pathogens was poorly described in the literature, despite their significance in livestock disease.

Optimise or design appropriate methodologies for the accurate identification and detection of clinically important Clostridia.

The phylogeny of 101 fully-sequenced clostridial species was inferred using 50 housekeeping genes, indicating taxonomic conflicts within the class (Chapter 3). By comparing distances between multilocus and individual gene phylogenetic trees, the *inf*B gene was identified as suitable genetic marker for the accurate identification of Clostridia to strain-resolution. The phylogenies inferred from this gene highlighted potentially misnamed strains and clinically-important pathogen subgroups than had already been acknowledged to date in the literature. Housekeeping gene sequences were again analysed to design a molecular diagnostic tool capable of simultaneously detecting key pathogenic species from soil samples. One protocol developed utilised automated multiplex ribosomal intergenic spacer analysis (AMRISA) to successfully detect and differentiate between clinically different pathogens from pure cultures, spiked soils and other environmental samples (Chapter 5). *In silico* analysis showed that the assay could distinguish between different strains and pathotypes of some pathogens. The protocol can allow a rapid, cost-effective method of screening environmental, food and clinical samples for a range of key clostridial pathogens.

Determine the influence of soil characteristics on the growth of key clostridial pathogens.

The pathogens *Cl. difficile, C. perfringens* and *C. sporogenes* persisted for >84-days when spiked into water-amended microcosm soils (Chapter 4), exhibiting two distinct periods of decay. All three species initially decayed rapidly, before transitioning into a slower stage of die-off. Die-off was significantly reduced, and pathogen abundance remained significantly higher in the drier soils compared to the wetter soils (see Chapter 4). The results suggest that drier conditions initiated spore production, thereby minimising overall population decline. There was no evidence of spiked or indigenous clostridial pathogen growth under any water treatment. The AMRISA protocol was used to identify clostridial pathogens from 540 field soils, and the change in relative abundance of pathogens between sites was determined (Chapter 6). Canonical correspondence analysis showed that short-term changes in relative soil moisture content had a small but statistically significant effect on changes in relative abundance with the clostridial community, while long-term waterlogging had a more profound effect. For some pathogens, relative abundance was significantly different between sites that were persistently waterlogged or dry, although the direction of change differed between species.

Assess the distribution and diversity of clostridial pathogens in agricultural soils.

Clostridia were detected virtually all of the agricultural soils examined (Chapter 6). Good quality peak profiles were visualised after AMRISA for over three-quarters of the samples, indicating widespread contamination across all sites. *C. novyi* and *Pn. sordellii* were putatively identified in circa 90% of samples. Other pathogens were also tentatively detected, including *C. baratii,* Group I *C. botulinum, C. chauvoei, C. perfringens* and *C. sporogenes.* Many of the peaks detected could not be identified, potentially indicating the presence of pathogens and strains genetically dissimilar to those currently sequenced. The total number and diversity of peaks (representing different clostridial strains) was lower during the coldest sampling period, indicating species die-off during cold periods.

Explore the impacts of farming regime on clostridial pathogens in the soil.

The change in pathogen prevalence and relative abundance between 4 different farms was explored using AMRISA (Chapter 6). Canonical correspondence analysis showed that differences in farm type explained the most variation in clostridial community of all the measured variables. Hurdle models indicated that for many detected species, presence/absence and relative abundance was significantly different between farm sites. Peaks affiliated with *Pn. sordellii* were more likely to be detected and had higher relative abundance under beef and dairy regimes. *C. perfringens* was more likely to be detected in beef and mixed grazing sites, and the relative abundance of *C. novyi* was significantly higher in mixed grazing regimes, compared to arable sites. Putative *C. chauvoei* and *C. baratii* peaks were also more likely to be detected with increased grazing. These results indicate that for some pathogens, the level of soil contamination is influenced by both the type and intensity of grazing regime.

Background clostridial pathogen distributions in soil are likely to be determined by underlying soil physicochemical parameters and the prevailing climate of the area. Populations fluctuate in response to seasonal changes, while this research indicates that agricultural practices also alter the prevalence and abundance of some pathogens in the soil, and the overall clostridial community composition. In combination with certain soil types and/or environmental conditions, farm practices such as intensive, year-on-year livestock grazing could create localised areas or certain time-periods when the risk of livestock infection is raised. To facilitate this work, a novel PCR-based AMRISA technique provided a suitable diagnostic for the simulatenous monitoring of many key clostridial pathogenic species in soils, as demonstrated, with further application recommended in other environmental, clinical and industrial settings.

CHAPTER 9: REFERENCES

Abascal, F., Zardoya, R. and Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38(suppl_2), pp.W7–W13.

Acosta-Martínez, V., Dowd, S., Sun, Y. and Allen, V., 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biology and Biochemistry*, [online] 40(11), pp.2762–2770. Available at: <http://www.sciencedirect.com/science/article/pii/S0038071708002484 > [Accessed 6 Oct. 2014].

Adak, G.K., 2002. Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000. *Gut*, [online] 51(6), pp.832–841. Available at: <http://gut.bmj.com/content/51/6/832> [Accessed 13 Oct. 2014].

Afshar, M., Raju, M., Ansell, D. and Bleck, T.P., 2011. Narrative review: tetanus—a health threat after natural disasters in developing countries. *Annals of Internal Medicine*, 154(5), pp.329–335.

Albrecht, J.A., 2005. *Clostridium perfringens*. [online] University of Nebraska- Food Saftey. Available at: <https://food.unl.edu/clostridium-perfringens> [Accessed 4 Jan. 2016].

Aldape, M.J., Bryant, A.E. and Stevens, D.L., 2006. *Clostridium sordellii* infection: epidemiology, clinical findings, and current perspectives on diagnosis and treatment. *Clinical Infectious Diseases*, [online] 43(11), pp.1436–46. Available at: <http://cid.oxfordjournals.org/content/43/11/1436.full> [Accessed 8 Apr. 2015].

Alfa, M.J., Robson, D., Davi, M., Bernard, K., Caeseele, P. Van and Harding, G.K.M., 2002. An outbreak of necrotizing enterocolitis associated with a novel *Clostridium* species in a neonatal intensive care unit. *Clinical Infectious Diseases*, 35(Supplement_1), pp.S101–S105.

Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, [online] 25(17), pp.3389–3402. Available at: <http://nar.oxfordjournals.org/content/25/17/3389.full> [Accessed 9 Jul. 2014].

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17), pp.3389–3402.

Andreesen, J.R., Bahl, H. and Gottschalk, G., 1989. Clostridia. In: N.P. Minton and D.J. Clarke, eds. [online] Boston, MA: Springer US, pp.27–62. Available at: <http://dx.doi.org/10.1007/978-1-4757-9718-3_2>.

Avery, L.M., Booth, P., Campbell, C., Tompkins, D. and Hough, R.L., 2012. Prevalence and survival of potential pathogens in source-segregated green waste compost. *Science of The Total Environment*, [online] 431, pp.128–138. Available at: <http://www.sciencedirect.com/science/article/pii/S0048969712006675>.

Bae, S. and Wuertz, S., 2009. Rapid decay of host-specific fecal Bacteroidales cells in seawater as measured by quantitative PCR with propidium monoazide. *Water Research*, 43(19), pp.4850–4859.

Bagge, E., Lewerin, S.S. and Johansson, K.E., 2009. Detection and identification by PCR of *Clostridium chauvoei* in clinical isolates, bovine faeces and substrates from biogas plant. *Acta Veterinaria Scandinavica*, 51(1), p.8.

Bagge, E., Persson, M. and Johansson, K.-E., 2010. Diversity of spore-forming bacteria in cattle manure, slaughterhouse waste and samples from biogas plants. *Journal of Applied Microbiology*, [online] 109(5), pp.1549–65. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20629803> [Accessed 10 Nov. 2014].

Bagge, E., Sahlström, L. and Albihn, A., 2005. The effect of hygienic treatment on the microbial flora of biowaste at biogas plants. *Water Research*, [online] 39, pp.4879–4886. Available at: <http://www.sciencedirect.com/science/article/pii/S0043135405001156>.

Bahl, H. and Dürre, P., 2001. *Clostridia: biotechnology and medical applications.* Wiley-VCH Verlag GmbH.

Bates, D., Maechler, M., Bolker, B. and Walker, S., 2014. *lme4: Linear mixed-effects models using Eigen and S4*. *R package version*, .

Baumgardner, D.J., 2012. Soil-related bacterial and fungal infections. *Journal of the American Board of Family Medicine : JABFM*, [online] 25(5), pp.734–44. Available at: <http://www.jabfm.org/content/25/5/734.short> [Accessed 23 Oct. 2014].

Båverud, V., 2002. *Clostridium difficile* infections in animals with special reference to the horse. A review. *The Veterinary Quarterly*, [online] 24(4), pp.203–19. Available at: <http://www.tandfonline.com/doi/abs/10.1080/01652176.2002.9695137#.VL1CQ9KsW5I> [Accessed 19 Jan. 2015].

Båverud, V., Gustafsson, A., Franklin, A., Aspán, A. and Gunnarsson, A., 2010. *Clostridium difficile*: prevalence in horses and environment, and antimicrobial susceptibility. *Equine Veterinary Journal*, [online] 35(5), pp.465–471. Available at: <http://doi.wiley.com/10.2746/042516403775600505> [Accessed 20 Oct. 2014].

Biswas, R., Raja, S., Sistla, S., Gopalakrishnan, M.S. and Saxena, S.K., 2018. Brain abscess and cervical lymphadenitis due to Paraclostridium bifermentans: A report of two cases. *Anaerobe*, 51, pp.8–11.

Blackwood, C.B., Hudleston, D., Zak, D.R. and Buyer, J.S., 2007. Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: insights from simulated microbial communities. *Applied and Environmental Microbiology*, 73(16), pp.5276–5283.

Böhnel, H., Briese, B.-H. and Gessler, F., 2002. Methods for health risk assessment by *Clostridium botulinum* in biocompost. In: *Microbiology of Composting*. Springer, pp.517–526.

Böhnel, H. and Lube, K., 2008. *Clostridium botulinum* and bio-compost. A contribution to the analysis of potential health hazards caused by bio-waste recycling. *Journal of Veterinary Medicine, Series B*, [online] 47(10), pp.785–795. Available at: <http://doi.wiley.com/10.1111/j.1439-0450.2000.00426.x> [Accessed 17 Nov. 2014].

Bonetta, S., Bonetta, S., Ferretti, E., Fezia, G., Gilli, G. and Carraro, E., 2014. Agricultural reuse of the digestate from anaerobic co-digestion of organic waste: microbiological contamination, metal hazards and fertilizing performance. *Water, Air, & Soil Pollution*, [online] 225(8), p.2046. Available at: <http://link.springer.com/10.1007/s11270-014-2046-2> [Accessed 11 Apr. 2016].

Boorman, D.B., Hollis, J.M. and Lilly, A., 1995. *Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom.* [online] NERC. Available at: <http://nora.nerc.ac.uk/id/eprint/7369>.

Boschiroli, M.L., Falkinham, J., Favre-Bonté, S., Nazaret, S., Piveteau, P., Sadowsky, M., Byappanahalli, M., Delaquis, P. and Hartmann, A., 2016. Natural soil reservoirs for human pathogenic and fecal indicator bacteria. In: *Manual of Environmental Microbiology, Fourth Edition*. American Society of Microbiology, p.3.

Le Bourhis, A.-G., Saunier, K., Doré, J., Carlier, J.-P., Chamba, J.-F., Popoff, M.-R. and Tholozan, J.-L., 2005. Development and validation of PCR primers to assess the diversity of *Clostridium* spp. in cheese by temporal temperature gradient gel electrophoresis. *Applied and Environmental Microbiology*, [online] 71(1), pp.29–38. Available at: <http://aem.asm.org/content/71/1/29.short> [Accessed 5 Jan. 2015].

Ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67(5), pp.1167–1179.

Ter Braak, C.J.F., 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, 69(1–3), pp.69–77.

ter Braak, C.J.F. and Šmilauer, P., 2015. Topics in constrained and unconstrained ordination. *Plant Ecology*, 216(5), pp.683–696.

Briolat, V. and Reysset, G., 2002. Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. *Journal of Bacteriology*, [online] 184(9), pp.2333–2343. Available at: <http://jb.asm.org/content/184/9/2333.abstract> [Accessed 4 Dec. 2014].

Brioukhanov, A.L. and Netrusov, A.I., 2007. Aerotolerance of strictly anaerobic microorganisms and factors of defense against oxidative stress: A review. *Applied Biochemistry and Microbiology*, [online] 43(6), pp.567–582. Available at: <http://link.springer.com/10.1134/S0003683807060014> [Accessed 4 Dec. 2014].

Brochier, V., Gourland, P., Kallassy, M., Poitrenaud, M. and Houot, S., 2012. Occurrence of pathogens in soils and plants in a long-term field study regularly amended with different composts and manure. *Agriculture, Ecosystems & Environment*, [online] 160, pp.91–98. Available at: <http://www.sciencedirect.com/science/article/pii/S0167880911001769> [Accessed 20 Oct. 2014].

Brocklehurst, T.F. and Lund, B.M., 1982. Isolation and properties of psychrotrophic and psychrophilic, pectolytic strains of *Clostridium*. *Journal of Applied Bacteriology*, [online] 53(3), pp.355–361. Available at: <http://doi.wiley.com/10.1111/j.1365-2672.1982.tb01283.x> [Accessed 29 Oct. 2014].

Brook, I., 2016. Spectrum and treatment of anaerobic infections. *Journal of Infection and Chemotherapy : Official Journal of the Japan Society of Chemotherapy*, [online] 22(1), pp.1–13. Available at: <http://www.sciencedirect.com/science/article/pii/S1341321X15002500> [Accessed 14 Mar. 2016].

Brouwer, A.F., Eisenberg, M.C., Remais, J. V, Collender, P.A., Meza, R. and Eisenberg, J.N.S., 2017. biexponen. *Environmental Science & Technology*, 51(4), pp.2186–2196.

Brüggemann, H., Bauer, R., Raffestin, S. and Gottschalk, G., 2004. Characterization of a heme oxygenase of *Clostridium tetani* and its

possible role in oxygen tolerance. *Archives of Microbiology*, [online] 182(2–3), pp.259–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15340779> [Accessed 4 Dec. 2014].

Bruggemann, H., Baumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A. and Gottschalk, G., 2003. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proceedings of the National Academy of Sciences*, [online] 100(3), pp.1316–21. Available at: <http://www.pnas.org/content/100/3/1316.long> [Accessed 5 Jun. 2015].

Brunt, J., Plowman, J., Gaskin, D., Itchner, M., Carter, A. and Peck, M., 2014. Functional characterisation of germinant receptors in *Clostridium botulinum* and *Clostridium sporogenes* presents novel insights into spore germination systems. *PLoS Pathogens*, [online] 10(9). Available at: <https://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=Y2kneCfSsV2zMNQjBkS&page=1&doc=2>.

Brynestad, S. and Granum, P.E., 2002. *Clostridium perfringens* and foodborne infections. *International Journal of Food Microbiology*, [online] 74(3), pp.195–202. Available at: <http://www.sciencedirect.com/science/article/pii/S0168160501006808> [Accessed 29 Sep. 2014].

Buchanan, J., 1993. *Struvite Control in Flush Water Recycle Components of Livestock Waste Management Systems*. University of Tennessee.

Bukovská, P., Jelínková, M., Hršelová, H., Sýkorová, Z. and Gryndler, M., 2010. Terminal restriction fragment length measurement errors are affected mainly by fragment length, G+ C nucleotide content and secondary structure melting point. *Journal of Microbiological Methods*, 82(3), pp.223–228.

Bulgari, D., Casati, P., Brusetti, L., Quaglino, F., Brasca, M., Daffonchio, D. and Bianco, P.A., 2009. Endophytic bacterial diversity in grapevine (Vitis vinifera L.) leaves described by 16S rRNA gene sequence analysis and length heterogeneity-PCR. *The Journal of Microbiology*, 47(4), pp.393–401.

Burke, C. and Darling, A.E., 2014. *Resolving microbial microdiversity with high accuracy full length 16S rRNA Illumina sequencing*. [online] *bioRxiv*. Cold Spring Harbor Labs Journals. Available at: <http://biorxiv.org/content/early/2014/11/06/010967.abstract> [Accessed 27 Jan. 2015].

Burlage, R.S. and Ellner, P.D., 1985. Comparison of the PRAS II, AN-Ident, and RapID-ANA systems for identification of anaerobic bacteria.

Journal of Clinical Microbiology, [online] 22(1), pp.32–35. Available at: <http://jcm.asm.org/content/22/1/32.short> [Accessed 22 Jan. 2015].

Burnham, K.P. and Anderson, D.R., 2003. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Burns, D.A., 2011. *Analysis of the spore germination mechanisms of Clostridium difficile*. [online] Available at: <http://etheses.nottingham.ac.uk/1852/1/David_Burns_PhD_Thesis.pdf > [Accessed 7 Oct. 2014].

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W. and Shipley, G.L., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4), pp.611–622.

Campbell, V., Legendre, P. and Lapointe, F.-J., 2009. Assessing congruence among ultrametric distance matrices. *Journal of Classification*, 26(1), p.103.

Cao, B., Li, R., Xiong, S., Yao, F., Liu, X., Wang, M., Feng, L. and Wang, L., 2011. Use of a DNA microarray for detection and identification of bacterial pathogens associated with fishery products. *Applied and Environmental Microbiology*, 77(23), pp.8219–8225.

Cao, X., Liu, X. and Dong, X., 2003. *Alkaliphilus crotonatoxidans* sp. nov., a strictly anaerobic, crotonate-dismutating bacterium isolated from a methanogenic environment. *International Journal of Systematic and Evolutionary Microbiology*, [online] 53(Pt 4), pp.971–5. Available at: <https://www.researchgate.net/publication/10635092_Alkaliphilus_crotonaoxidans_sp_nov_a_strictly_anaerobic_crotonate-dismutating_bacterium_isolated_from_a_methanogenic_environment> [Accessed 24 Feb. 2016].

Carlin, F., 2011. Origin of bacterial spores contaminating foods. *Food Microbiology*, [online] 28(2), pp.177–82. Available at: <http://www.sciencedirect.com/science/article/pii/S0740002010001838 > [Accessed 26 Sep. 2014].

Cherif, H., Ouzari, H., Marzorati, M., Brusetti, L., Jedidi, N., Hassen, A. and Daffonchio, D., 2008. Bacterial community diversity assessment in municipal solid waste compost amended soil using DGGE and ARISA fingerprinting methods. *World Journal of Microbiology and Biotechnology*, 24(7), pp.1159–1167.

Chessbrough, M., 2002. *District Laboratory Practice in Tropical Countries. Part 2*. *Cambridge University Press. Campbridge CB2*, Cambridge.

Chin, K.-J., Rainey, F.A., Janssen, P.H. and Conrad, R., 1998.

Methanogenic degradation of polysaccharides and the characterization of polysaccharolytic clostridia from anoxic rice field soil. *Systematic and Applied Microbiology*, 21(2), pp.185–200.

Chowdhury, P.R., DeMaere, M., Chapman, T., Worden, P., Charles, I.G., Darling, A.E. and Djordjevic, S.P., 2016. Comparative genomic analysis of toxin-negative strains of *Clostridium difficile* from humans and animals with symptoms of gastrointestinal disease. *BMC Microbiology*, 16(1), p.41.

Christensen, H., Kuhnert, P., Olsen, J.E. and Bisgaard, M., 2004. Comparative phylogenies of the housekeeping genes atpD, infB and rpoB and the 16S rRNA gene within the Pasteurellaceae. *International Journal of Systematic and Evolutionary Microbiology*, [online] 54(Pt 5), pp.1601–9. Available at: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.03018-0> [Accessed 26 Jan. 2016].

Ciani, A., Goss, K.-U. and Schwarzenbach, R.P., 2005. Light penetration in soil and particulate minerals. *European Journal of Soil Science*, [online] 56(5), pp.561–574. Available at: <http://doi.wiley.com/10.1111/j.1365-2389.2005.00688.x> [Accessed 4 Dec. 2014].

Cilia, V., Lafay, B. and Christen, R., 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Molecular Biology and Evolution*, [online] 13(3), pp.451–61. Available at: <http://mbe.oxfordjournals.org/content/13/3/451.abstract> [Accessed 16 Feb. 2016].

College, S.A., 2012. Further outbreaks of blackleg seen across Scotland. *Veterinary Record*.

Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. and Farrow, J.A.E., 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *International Journal of Systematic Bacteriology*, [online] 44(4), pp.812–826. Available at: <http://ijs.sgmjournals.org/content/44/4/812.short> [Accessed 13 Oct. 2014].

Consortium, T.U., 2011. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Research*, [online] 39(Database issue), pp.D214-9. Available at: <http://nar.oxfordjournals.org/content/39/suppl_1/D214.full> [Accessed 26 Aug. 2015].

Cordoba, M., Aranda, E., Medina, L., Jordano, R. and Cordoba, J., 2001. Differentiation of *Clostridium perfringens* and *Clostridium botulinum* from non-toxigenic clostridia, isolated from prepared and frozen foods by PCR-

DAN based methods. *Nahrung-Food*, [online] 45(2), pp.125–128. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=U1wE1Kv2WWqS6aotfwQ&page=1&doc=1> [Accessed 4 Nov. 2014].

Cotter, J.L., Chinn, M.S. and Grunden, A.M., 2009. Influence of process parameters on growth of *Clostridium ljungdahlii* and *Clostridium autoethanogenum* on synthesis gas. *Enzyme and Microbial Technology*, [online] 44(5), pp.281–288. Available at: <http://www.sciencedirect.com/science/article/pii/S0141022908002895> [Accessed 24 Feb. 2016].

Crane, S.R. and Moore, J.A., 1986. Modeling enteric bacterial die-off: A review. *Water, Air, and Soil Pollution*, [online] 27(3), pp.411–439. Available at: <https://doi.org/10.1007/BF00649422>.

Creti, R., Fenicia, L. and Aureli, P., 1990. Occurrence of *Clostridium botulinum* in the soil of the vicinity of Rome. *Current Microbiology*, [online] 20(5), pp.317–321. Available at: <http://link.springer.com/10.1007/BF02091912> [Accessed 20 Oct. 2014].

Culman, S.W., Gauch, H.G., Blackwood, C.B. and Thies, J.E., 2008. Analysis of T-RFLP data using analysis of variance and ordination methods: a comparative study. *Journal of microbiological methods*, [online] 75(1), pp.55–63. Available at: <http://www.sciencedirect.com/science/article/pii/S0167701208001723> [Accessed 2 Sep. 2014].

Dahllöf, I., 2002. Molecular community analysis of microbial diversity. *Current Opinion in Biotechnology*, [online] 13(3), pp.213–217. Available at: <http://www.sciencedirect.com/science/article/pii/S0958166902003142> [Accessed 15 Oct. 2014].

Daubin, V., Gouy, M. and Perrière, G., 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Research*, [online] 12(7), pp.1080–90. Available at: <http://genome.cshlp.org/content/12/7/1080.short> [Accessed 12 Feb. 2016].

Davies, C.M., Long, J.A., Donald, M. and Ashbolt, N.J., 1995. Survival of fecal microorganisms in marine and freshwater sediments. *Applied and Environmental Microbiology*, [online] 61(5), pp.1888–96. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7646026> [Accessed 8 Apr. 2016].

Dezfulian, M., 1999. Aerobic growth and toxigenicity of *Clostridium botulinum* types A and B. *Folia Microbiologica*, [online] 44(2), pp.167–

170. Available at: <http://link.springer.com/10.1007/BF02816236> [Accessed 10 Dec. 2014].

Dhalluin, A., Lemée, L., Pestel-Caron, M., Mory, F., Leluan, G., Lemeland, J.-F. and Pons, J.-L., 2003. Genotypic differentiation of twelve *Clostridium* species by polymorphism analysis of the triosephosphate isomerase (tpi) gene. *Systematic and applied microbiology*, [online] 26(1), pp.90–6. Available at: <http://www.sciencedirect.com/science/article/pii/S0723202004701642> [Accessed 7 May 2015].

Dineen, S.M., Aranda, R., Anders, D.L. and Robertson, J.M., 2010. An evaluation of commercial DNA extraction kits for the isolation of bacterial spore DNA from soil. *Journal of Applied Microbiology*, 109(6), pp.1886–1896.

Dingle, K.E., Griffiths, D., Didelot, X., Evans, J., Vaughan, A., Kachrimanidou, M., Stoesser, N., Jolley, K.A., Golubchik, T., Harding, R.M., Peto, T.E., Fawley, W., Walker, A.S., Wilcox, M. and Crook, D.W., 2011. Clinical *Clostridium difficile*: clonality and pathogenicity locus diversity. *PloS one*, [online] 6(5), p.e19993. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0019993> [Accessed 24 Feb. 2016].

Dodds, K.L., 1992. *Clostridium botulinum* in the environment. In: *Clostridium botulinum: Ecology and Control in Foods*. New York: CRC Press, p.21.

Donskey, C.J., Sunkesula, V.C.K., Jencson, A.L., Stone, N.D., Gould, C. V, McDonald, L.C., Samore, M., Mayer, J., Pacheco, S., Sambol, S., Petrella, L., Terry, D. and Gerding, D.N., 2014. Utility of a commercial PCR assay and a clinical prediction rule for detection of toxigenic *Clostridium difficile* in asymptomatic carriers. *Journal of Clinical Microbiology*, [online] 52(1), pp.315–8. Available at: <http://jcm.asm.org/content/52/1/315.short> [Accessed 16 Jan. 2015].

Dorr de Quadros, P., Zhalnina, K., Davis-Richardson, A., Fagen, J.R., Drew, J., Bayer, C., Camargo, F.A.O. and Triplett, E.W., 2012. The effect of tillage system and crop rotation on soil microbial diversity and composition in a subtropical acrisol. *Diversity*, [online] 4(4), pp.375–395. Available at: <http://www.mdpi.com/1424-2818/4/4/375/htm> [Accessed 6 Nov. 2014].

Dudley, D.J., Guentzel, M.N., Ibarra, M.J., Moore, B.E. and Sagik, B.P., 1980. Enumeration of potentially pathogenic bacteria from sewage sludges. *Applied and Environmental Microbiology*, [online] 39(1), pp.118–126. Available at: <http://aem.asm.org/content/39/1/118.abstract> [Accessed 20 Nov. 2014].

Dunbar, J., Ticknor, L.O. and Kuske, C.R., 2000. Assessment of microbial

diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. *Applied and Environmental Microbiology*, 66(7), pp.2943–2950.

Dürre, P., 2004. *Handbook on clostridia*. CRC Press.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, [online] 32(5), pp.1792–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=390337&tool=pmcentrez&rendertype=abstract> [Accessed 11 Jul. 2014].

Edwards, A.N., Suárez, J.M. and McBride, S.M., 2013. Culturing and maintaining *Clostridium difficile* in an anaerobic environment. *Journal of Visualized Experiments*, [online] (79), p.e50787. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3871928&tool=pmcentrez&rendertype=abstract> [Accessed 22 Jul. 2015].

Elliott, G.N., Thomas, N., MacRae, M., Campbell, C.D., Ogden, I.D. and Singh, B.K., 2012. Multiplex T-RFLP Allows for Increased Target Number and Specificity: Detection of Salmonella enterica and Six Species of Listeria in a Single Test. *PloS one*, 7, p.e43672.

van Elsas, J.D., Trevors, J.T. and van Overbeek, L.S., 1991. Influence of soil properties on the vertical movement of genetically-marked Pseudomonas fluorescens through large soil microcosms. *Biology and Fertility of Soils*, [online] 10(4), pp.249–255. Available at: <http://link.springer.com/10.1007/BF00337375> [Accessed 28 Oct. 2014].

England, P.H., 2016. *The characteristics, diagnosis and epidemiology of Clostridium perfringens*. [online] Available at: <https://www.gov.uk/guidance/clostridium-perfringens>.

Espelund, M. and Klaveness, D., 2014. Botulism outbreaks in natural environments - an update. *Frontiers in Microbiology*, [online] 5(June), p.287. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4052663&tool=pmcentrez&rendertype=abstract> [Accessed 23 Oct. 2014].

Euzéby, J.P., 1997. List of bacterial names with standing in nomenclature: a folder available on the Internet. *International Journal of Systematic and Evolutionary Microbiology*, 47(2), pp.590–592.

Evans, R.I., Russell, N.J., Gould, G.W. and McClure, P.J., 1997. The germinability of spores of a psychrotolerant, non-proteolytic strain of *Clostridium botulinum* is influenced by their formation and storage temperature. *Journal of Applied Microbiology*, [online] 83(3), pp.273–280. Available at: <http://doi.wiley.com/10.1046/j.1365-2672.1997.00225.x> [Accessed 4 Nov. 2014].

Fach, P., Fenicia, L., Knutsson, R., Wielinga, P.R., Anniballi, F., Delibato, E., Auricchio, B., Woudstra, C., Agren, J., Segerman, B., de Medici, D. and van Rotterdam, B.J., 2011. An innovative molecular detection tool for tracking and tracing *Clostridium botulinum* types A, B, E, F and other botulinum neurotoxin producing Clostridia based on the GeneDisc cycler. *International Journal of Food Microbiology*, [online] 145 Suppl, pp.S145-51. Available at: <http://www.sciencedirect.com/science/article/pii/S0168160510001911> [Accessed 5 Aug. 2015].

Fach, P., Micheau, P., Mazuet, C., Perelle, S. and Popoff, M., 2009. Development of real-time PCR tests for detecting botulinum neurotoxins A, B, E, F producing *Clostridium botulinum, Clostridium baratii* and *Clostridium butyricum*. *Journal of Applied Microbiology*, [online] 107(2), pp.465–73. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19291235> [Accessed 5 Nov. 2014].

Fang, F.C., Polage, C.R. and Wilcox, M.H., 2017. Point-Counterpoint: What is the optimal approach for detection of *Clostridium difficile* infection? *Journal of Clinical Microbiology*, p.JCM-02463.

Feligini, M., Panelli, S., Sacchi, R., Ghitti, M. and Capelli, E., 2015. Tracing the origin of raw milk from farm by using Automated Ribosomal Intergenic Spacer Analysis (ARISA) fingerprinting of microbiota. *Food Control*, 50, pp.51–56.

Fenicia, L., Fach, P., van Rotterdam, B.J., Anniballi, F., Segerman, B., Auricchio, B., Delibato, E., Hamidjaja, R.A., Wielinga, P.R., Woudstra, C., Agren, J., De Medici, D. and Knutsson, R., 2011. Towards an international standard for detection and typing botulinum neurotoxin-producing Clostridia types A, B, E and F in food, feed and environmental samples: a European ring trial study to evaluate a real-time PCR assay. *International journal of food microbiology*, [online] 145 Suppl, pp.S152-7. Available at: <http://www.sciencedirect.com/science/article/pii/S0168160511000687> [Accessed 16 Jan. 2015].

Finegold, S.M., Song, Y. and Liu, C., 2002. Taxonomy--General comments and update on taxonomy of Clostridia and Anaerobic cocci. *Anaerobe*, [online] 8(5), pp.283–5. Available at: <http://www.sciencedirect.com/science/article/pii/S1075996402904361> [Accessed 20 Aug. 2015].

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E. and Butler, G., 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evolutionary Biology*, [online] 6(1), p.99. Available at: <http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-6-99> [Accessed 12 Feb. 2016].

Forsberg, K.J., Reyes, A., Wang, B., Selleck, E.M., Sommer, M.O.A. and Dantas, G., 2012. The shared antibiotic resistome of soil bacteria and human pathogens. *Science*, 337(6098), pp.1107–1111.

Frostegård, Å., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., Nesme, X. and Simonet, P., 1999. Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology*, 65(12), pp.5409–5420.

Fu, Y.-X. and Li, W.-H., 1993. Statistical tests of neutrality of mutations. *Genetics*, 133(3), pp.693–709.

Gadagkar, S.R., Rosenberg, M.S. and Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, [online] 304(1), pp.64–74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15593277> [Accessed 18 Jan. 2016].

Garbeva, P., van Veen, J.A. and van Elsas, J.D., 2004. Microbial diversity in soil: selection microbial populations by plant and soil type and implications for disease suppressiveness. *Annual Review of Phytopathology*, [online] 42, pp.243–70. Available at: <http://www.annualreviews.org/doi/abs/10.1146/annurev.phyto.42.012604.135455> [Accessed 7 Oct. 2014].

Garcia, M.M. and McKay, K.A., 1969. On the growth and survival of *Clostridium septicum* in soil. *Journal of Applied Bacteriology*, [online] 32(3), pp.362–370. Available at: <http://doi.wiley.com/10.1111/j.1365-2672.1969.tb00985.x> [Accessed 9 Oct. 2014].

Garrec, N., Picard-Bonnaud, F. and Pourcher, A.., 2003. Occurrence of *Listeria sp.* and *L. monocytogenes* in sewage sludge used for land application: effect of dewatering, liming and storage in tank on survival of *Listeria* species. *FEMS Immunology & Medical Microbiology*, [online] 35(3), pp.275–283. Available at: <http://www.sciencedirect.com/science/article/pii/S0928824402004431> [Accessed 30 Oct. 2014].

Gehr, R., Wagner, M., Veerasubramanian, P. and Payment, P., 2003. Disinfection efficiency of peracetic acid, UV and ozone after enhanced primary treatment of municipal wastewater. *Water Research*, [online] 37(19), pp.4573–86. Available at: <http://www.sciencedirect.com/science/article/pii/S0043135403003944> [Accessed 1 Nov. 2014].

Gessler, F. and Böhnel, H., 2006. Persistence and mobility of a *Clostridium botulinum* spore population introduced to soil with spiked compost. *FEMS Microbiology Ecology*, [online] 58(3), pp.384–93. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17117983>

[Accessed 21 Oct. 2014].

Ghodbane, R., Medie, F.M., Lepidi, H., Nappez, C. and Drancourt, M., 2014. Long-term survival of tuberculosis complex mycobacteria in soil. *Microbiology*, 160(3), pp.496–501.

Ghoddusi, H.B., Sherburn, R.E. and Aboaba, O.O., 2013. Growth Limiting pH, Water Activity, and Temperature for Neurotoxigenic Strains of *Clostridium butyricum*. *International Scholarly Research Notices*, 2013.

Gil, R., Silva, F.J., Peretó, J. and Moya, A., 2004. Determination of the core of a minimal bacterial gene set. *Microbiology and molecular biology reviews : MMBR*, [online] 68(3), p.518–37, table of contents. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=515251&tool=pmcentrez&rendertype=abstract> [Accessed 17 Mar. 2015].

Gilboa, Y. and Friedler, E., 2008. UV disinfection of RBC-treated light greywater effluent: kinetics, survival and regrowth of selected microorganisms. *Water Research*, [online] 42(4–5), pp.1043–50. Available at: <http://www.sciencedirect.com/science/article/pii/S0043135407006306> [Accessed 2 Dec. 2014].

Girardin, H., Morris, C.E., Albagnac, C., Dreux, N., Glaux, C. and Nguyen-The, C., 2005. Behaviour of the pathogen surrogates *Listeria* innocua and *Clostridium sporogenes* during production of parsley in fields fertilized with contaminated amendments. *FEMS microbiology ecology*, [online] 54(2), pp.287–95. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16332327> [Accessed 7 Oct. 2014].

Goh, S., Riley, T. V and Chang, B.J., 2005. Isolation and characterization of temperate bacteriophages of *Clostridium difficile*. *Applied and Environmental Microbiology*, [online] 71(2), pp.1079–83. Available at: <http://aem.asm.org/content/71/2/1079.full> [Accessed 19 Dec. 2014].

Grecz, N. and Arvay, L.H., 1982. Effect of temperature on spore germination and vegetative cell growth of *Clostridium botulinum*. *Applied and Environmental Microbiology*, [online] 43(2), pp.331–337. Available at: <http://aem.asm.org/content/43/2/331.short>.

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D.W., Fung, R., Golubchik, T., Harding, R.M., Jeffery, K.J.M., Jolley, K.A., Kirton, R., Peto, T.E., Rees, G., Stoesser, N., Vaughan, A., Walker, A.S., Young, B.C., Wilcox, M. and Dingle, K.E., 2010. Multilocus sequence typing of *Clostridium difficile. Journal of Clinical Microbiology*, [online] 48(3), pp.770–8. Available at: <http://jcm.asm.org/content/48/3/770.short> [Accessed 3 Mar. 2016].

Grimes, D.J., 1991. Ecology of estuarine bacteria capable of causing human disease: a review. *Estuaries*, [online] 14(4), p.345. Available at: <http://link.springer.com/10.2307/1352260> [Accessed 12 Apr. 2016].

Gupta, R.S. and Gao, B., 2009. Phylogenomic analyses of clostridia and identification of novel protein signatures that are specific to the genus *Clostridium sensu stricto* (cluster I). *International Journal of Systematic and Evolutionary Microbiology*, [online] 59(Pt 2), pp.285–94. Available at: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.001792-0#tab2> [Accessed 12 Feb. 2016].

Gurtler, V., Wilson, V.A. and Mayall, B.C., 1991. Classification of medically important clostridia using restriction endonuclease site differences of PCR-amplified 16S rDNA. *Journal of General Microbiology*, [online] 137(11), pp.2673–2679. Available at: <http://mic.sgmjournals.org/content/137/11/2673.short> [Accessed 6 Oct. 2014].

Haagsma, J., 1991. Pathogenic anaerobic bacteria and the environment. *Revue Scientifique Et Technique*, 10, pp.749–764.

Halm, A., Wagner, M., Köfer, J. and Hein, I., 2010. Novel real-time PCR assay for simultaneous detection and differentiation of *Clostridium chauvoei* and *Clostridium septicum* in clostridial myonecrosis. *Journal of Clinical Microbiology*, 48(4), pp.1093–1098.

Hang'ombe, B.M., Isogai, E., Lungu, J., Mubita, C., Nambota, A., Kirisawa, R., Kimura, K. and Isogai, H., 2000. Detection and characterization of *Clostridium* species in soil of Zambia. *Comparative Immunology, Microbiology and Infectious Diseases*, [online] 23(4), pp.277–284. Available at: <http://www.sciencedirect.com/science/article/pii/S0147957199000788> [Accessed 20 Oct. 2014].

Hargreaves, K.R., Colvin, H. V, Patel, K. V, Clokie, J.J.P. and Clokie, M.R.J., 2013. Genetically diverse *Clostridium difficile* strains harboring abundant prophages in an estuarine environment. *Applied and Environmental Microbiology*, [online] 79(20), pp.6236–43. Available at: <http://aem.asm.org/content/79/20/6236.full> [Accessed 19 Dec. 2014].

Harrell Jr, F.E. and Harrell Jr, M.F.E., 2018. Package 'Hmisc'. *R Foundation for Statistical Computing*.

Harrison, C.J. and Langdale, J.A., 2006. A step by step guide to phylogeny reconstruction. *The Plant journal : for Cell and Molecular Biology*, [online] 45(4), pp.561–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16441349> [Accessed 9 Mar. 2015].

Harrison, X.A., 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, 2, p.e616.

Hatheway, C.L., 1990. Toxigenic clostridia. *Clinical Microbiology Reviews*, [online] 3(1), pp.66–98. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=358141&tool=pmcentrez&rendertype=abstract> [Accessed 6 Oct. 2014].

Hayashi, H., Sakamoto, M., Kitahara, M. and Benno, Y., 2003. Molecular analysis of fecal microbiota in elderly individuals using 16S rDNA library and T-RFLP. *Microbiology and Immunology*, [online] 47(8), pp.557–570. Available at: <http://doi.wiley.com/10.1111/j.1348-0421.2003.tb03418.x> [Accessed 6 Oct. 2014].

He, L., Chen, W. and Liu, Y., 2006. Production and partial characterization of bacteriocin-like pepitdes by *Bacillus licheniformis* ZJU12. *Microbiological Research*, [online] 161(4), pp.321–6. Available at: <http://www.sciencedirect.com/science/article/pii/S0944501305001229> [Accessed 19 Dec. 2014].

Head, C.B. and Ratnam, S., 1988. Comparison of API ZYM system with API AN-Ident, API 20A, Minitek Anaerobe II, and RapID-ANA systems for identification of *Clostridium difficile. Journal of Clinical Microbiology*, [online] 26(1), pp.144–146. Available at: <http://jcm.asm.org/content/26/1/144.short> [Accessed 22 Jan. 2015].

Heinänen, S., Rönkä, M. and Von Numers, M., 2008. Modelling the occurrence and abundance of a colonial species, the arctic tern *Sterna paradisaea* in the archipelago of SW Finland. *Ecography*, 31(5), pp.601–611.

Henegariu, O., Heerema, N.A., Dlouhy, S.R., Vance, G.H. and Vogt, P.H., 1997. Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, 23(3), pp.504–511.

Hengstmann, U., Chin, K.-J., Janssen, P.H. and Liesack, W., 1999. Comparative phylogenetic assignment of environmental sequences of genes encoding 16S rRNA and numerically abundant culturable bacteria from an anoxic rice paddy soil. *Applied and Environmental Microbiology*, 65(11), pp.5050–5058.

Hielm, S., Hyytiä, E., Ridell, J. and Korkeala, H., 1996. Detection of *Clostridium botulinum* in fish and environmental samples using polymerase chain reaction. *International Journal of Food Microbiology*, [online] 31(1–3), pp.357–365. Available at: <http://www.sciencedirect.com/science/article/pii/0168160596009841> [Accessed 7 Oct. 2014].

Hijnen, W.A.M., Beerendonk, E.F. and Medema, G.J., 2006. Inactivation credit of UV radiation for viruses, bacteria and protozoan (oo)cysts in

water: a review. *Water Research*, [online] 40(1), pp.3–22. Available at: <http://www.sciencedirect.com/science/article/pii/S004313540500610X> [Accessed 26 Nov. 2014].

Hill, C.R., Butler, J.M. and Vallone, P.M., 2009. A 26plex autosomal STR assay to aid human identity testing. *Journal of Forensic Sciences*, 54(5), pp.1008–1015.

Hill, G.B. and Osterhout, S., 1972. Experimental Effects of Hyperbaric Oxygen on Selected Clostridial Species. I. In-Vitro Studies. *Journal of Infectious Diseases*, [online] 125(1), pp.17–25. Available at: <http://jid.oxfordjournals.org/content/125/1/17.short> [Accessed 3 Jan. 2015].

Hothorn, T., Zeileis, A., Farebrother, R.W., Cummins, C., Millo, G. and Mitchell, D., 2015. *lmtest: Testing linear regression models*. *R package version 0.9-34*.

Huang, S.-W., Chan, J.P.-W., Shia, W.-Y., Shyu, C.-L., Tung, K.-C. and Wang, C.-Y., 2013a. The utilization of a commercial soil nucleic acid extraction kit and PCR for the detection of *Clostridium tetanus* and *Clostridium chauvoei* on farms after flooding in Taiwan. *Journal of Veterinary Medical Science*, [online] 75(4), pp.489–495. Available at: <https://www.jstage.jst.go.jp/article/jvms/75/4/75_12-0271/_article> [Accessed 23 Mar. 2016].

Huang, Y., Sun, Y., Ma, S., Chen, L., Zhang, H. and Deng, Y., 2013b. Isolation and characterization of *Keratinibaculum paraultunense* gen. nov., sp. nov., a novel thermophilic, anaerobic bacterium with keratinolytic activity. *FEMS Microbiology Letters*, [online] 345(1), pp.56–63. Available at: <http://femsle.oxfordjournals.org/content/345/1/56.abstract> [Accessed 2 Mar. 2016].

Hughes, C., Gillespie, I.A. and O'Brien, S.J., 2007. Foodborne transmission of infectious intestinal disease in England and Wales, 1992–2003. *Food Control*, [online] 18(7), pp.766–772. Available at: <http://www.sciencedirect.com/science/article/pii/S0956713506000387> [Accessed 13 Oct. 2014].

Hung, C., Cheng, C., Cheng, L., Liang, C. and Lin, C., 2008. Application of *Clostridium*-specific PCR primers on the analysis of dark fermentation hydrogen-producing bacterial community. *International Journal of Hydrogen Energy*, [online] 33(5), pp.1586–1592. Available at: <http://www.sciencedirect.com/science/article/pii/S036031990700568X> [Accessed 6 Apr. 2015].

Hunter, L.C., Miller, J.K. and Poxton, I.R., 1999. The association of *Clostridium botulinum* type C with equine grass sickness: a toxicoinfection? *Equine Veterinary Journal*, [online] 31(6), pp.492–499.

Available at: <http://doi.wiley.com/10.1111/j.2042-3306.1999.tb03857.x> [Accessed 14 Oct. 2014].

Huss, H.H., 1980. Distribution of *Clostridium botulinum*. *Applied and Environmental Microbiology*, [online] 39(4), pp.764–769. Available at: <http://aem.asm.org/content/39/4/764.short> [Accessed 7 Oct. 2014].

Ibekwe, A.M., Kennedy, A.C., Frohne, P.S., Papiernik, S.K., Yang, C.-H. and Crowley, D.E., 2002. Microbial diversity along a transect of agronomic zones. *FEMS Microbiology Ecology*, [online] 39(3), pp.183–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19709198> [Accessed 16 Oct. 2014].

Van Immerseel, F., De Buck, J., Pasmans, F., Huyghebaert, G., Haesebrouck, F. and Ducatelle, R., 2004. *Clostridium perfringens* in poultry: an emerging threat for animal and public health. *Avian Pathology*, [online] 33(6), pp.537–49. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=R2LgFtlEhlmF8uBh2MG&page=2&doc=17> [Accessed 6 Oct. 2014].

Jackman, S., Fearon, J., Jackman, M.S. and MCMCpack, S., 2007. The pscl package. *Soware. http://cran. rproject. org/src/contrib/Descriptions/pscl. html*.

Jamieson, R.C., Gordon, R.J., Sharples, K.E., Stratton, G.W. and Madani, A., 2002. Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review. *Canadian Biosystems Engineering*, 44(1), pp.1–9.

Janda, J.M. and Abbott, S.L., 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9), pp.2761–2764.

Janssen, P.H., 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology*, [online] 72(3), pp.1719–28. Available at: <http://aem.asm.org/content/72/3/1719.short> [Accessed 24 Sep. 2014].

Janvilisri, T., Scaria, J., Gleed, R., Fubini, S., Bonkosky, M.M., Gröhn, Y.T. and Chang, Y.-F., 2010. Development of a microarray for identification of pathogenic *Clostridium* spp. *Diagnostic Microbiology and Infectious Disease*, [online] 66(2), pp.140–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2815050&tool=pmcentrez&rendertype=abstract> [Accessed 8 Apr. 2015].

Jean, D., Briolat, V. and Reysset, G., 2004. Oxidative stress response in *Clostridium perfringens*. *Microbiology*, [online] 150(Pt 6), pp.1649–59. Available at: <http://mic.sgmjournals.org/content/150/6/1649.short>

[Accessed 10 Dec. 2014].

Jeffery, S. and Van der Putten, W.H., 2011. *Soil borne human diseases*. [online] Available at: <https://www.kiza.beroepsziekten.nl/sites/default/files/documenten/soil borne human diseases jeffery putte 2011.pdf>.

Jensen, M., Genigeorgis, C. and Lindroth, S., 1987. Probability of growth of *Clostridium botulinum* as affected by strain, cell and serologic type, inoculum size and temperature and time of incubation in a model broth system. *Journal of Food Safety*, 8(2), pp.109–126.

Jordan, I.K., Rogozin, I.B., Wolf, Y.I. and Koonin, E. V., 2002. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Research*, [online] 12(6), pp.962–968. Available at: <http://genome.cshlp.org/content/12/6/962.full#cited-by> [Accessed 20 May 2015].

Kalendar, R., Khassenov, B., Ramankulov, Y., Samuilova, O. and Ivanov, K.I., 2017. FastPCR: An in silico tool for fast primer and probe design and advanced sequence analysis. *Genomics*.

Kaplan, C.W. and Kitts, C.L., 2003. Variation between observed and true terminal restriction fragment length is dependent on true TRF length and purine content. *Journal of Microbiological Methods*, 54(1), pp.121–125.

Kenyon, F., Sargison, N.D., Skuce, P.J. and Jackson, F., 2009. Sheep helminth parasitic disease in south eastern Scotland arising as a possible consequence of climate change. *Veterinary Parasitology*, 163(4), pp.293–297.

Keto-Timonen, R., Heikinheimo, A., Eerola, E. and Korkeala, H., 2006. Identification of Clostridium species and DNA fingerprinting of *Clostridium perfringens* by amplified fragment length polymorphism analysis. *Journal of Clinical Microbiology*, [online] 44(11), pp.4057–65. Available at: <http://jcm.asm.org/content/44/11/4057.short> [Accessed 13 Oct. 2014].

Khan, I.U.H., Gannon, V., Kent, R., Koning, W., Lapen, D.R., Miller, J., Neumann, N., Phillips, R., Robertson, W., Topp, E., van Bochove, E. and Edge, T.A., 2007. Development of a rapid quantitative PCR assay for direct detection and quantification of culturable and non-culturable *Escherichia coli* from agriculture watersheds. *Journal of Microbiological Methods*, [online] 69(3), pp.480–8. Available at: <http://www.sciencedirect.com/science/article/pii/S0167701207000693> [Accessed 6 Jan. 2015].

Khanna, S., Goyal, A. and Moholkar, V.S., 2013. Production of n-butanol from biodiesel derived crude glycerol using *Clostridium pasteurianum* immobilized on Amberlite. *Fuel*, 112, pp.557–561.

Khoruts, A., Dicksved, J., Jansson, J.K. and Sadowsky, M.J., 2010. Changes in the composition of the human fecal microbiome following bacteriotherapy for recurrent *Clostridium difficile*-associated diarrhea. *Journal of Clinical Gastroenterology*, [online] 44(5), pp.354–360. Available at: <http://escholarship.org/uc/item/0582t44m> [Accessed 6 Oct. 2014].

Kibbe, W.A., 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, 35(suppl_2), pp.W43–W46.

Kim, J., Lee, D., Lee, K., Choi, C. and Kang, K., 2004. Distribution and antimicrobial susceptibility of *Clostridium* species in soil contaminated with domestic livestock feces of Korea. *Journal of Microbiology and Biotechnology*, [online] 14(2), pp.401–410. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=T1mgMPRNUmx9cSpSHgB&page=2&doc=55> [Accessed 30 Oct. 2014].

Kirk, J.L., Beaudette, L.A., Hart, M., Moutoglis, P., Klironomos, J.N., Lee, H. and Trevors, J.T., 2004. Methods of studying soil microbial diversity. *Journal of microbiological methods*, [online] 58(2), pp.169–88. Available at: <http://www.sciencedirect.com/science/article/pii/S0167701204000983> [Accessed 10 Jul. 2014].

Klobutcher, L.A., Ragkousi, K. and Setlow, P., 2006. The *Bacillus subtilis* spore coat provides "eat resistance" during phagocytic predation by the protozoan Tetrahymena thermophila. *Proceedings of the National Academy of Sciences*, 103(1), pp.165–170.

Kolek, J., Patáková, P., Melzoch, K., Sigler, K. and Řezanka, T., 2015. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as determined by lipidomic analysis. *PloS one*, [online] 10(3), p.e0122058. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122058#pone.0122058.ref003> [Accessed 26 Feb. 2016].

Köser, C.U., Ellington, M.J., Cartwright, E.J.P., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T.G., Dougan, G., Bentley, S.D. and Parkhill, J., 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathogens*, 8(8), p.e1002824.

Kuhner, M.K. and Yamato, J., 2014. Practical performance of tree comparison metrics. *Systematic Biology*, 64(2), pp.205–214.

Kuhnert, P., Capaul, S.E., Nicolet, J. and Frey, J., 1996. Phylogenetic positions of C*lostridium chauvoei* and *Clostridium septicum* based on 16S rRNA gene sequences. *International Journal of Systematic Bacteriology*, [online] 46(4), pp.1174–1176. Available at:

<http://ijs.sgmjournals.org/content/46/4/1174.short> [Accessed 14 Oct. 2014].

Kuske, C.R., Barns, S.M., Grow, C.C., Merrill, L. and Dunbar, J., 2006. Environmental survey for four pathogenic bacteria and closely related species using phylogenetic and functional genes. *Journal of Forensic Sciences*, [online] 51(3), pp.548–58. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16696701> [Accessed 7 Oct. 2014].

Lanao, M., Ormad, M.P., Goñi, P., Miguel, N., Mosteo, R. and Ovelleiro, J.L., 2010. Inactivation of C*lostridium perfringens* spores and vegetative cells by photolysis and TiO2 photocatalysis with H2O2. *Solar Energy*, [online] 84(4), pp.703–709. Available at: <http://www.sciencedirect.com/science/article/pii/S0038092X10000368> [Accessed 5 Nov. 2014].

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. and Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), pp.772–773.

Lang, J.M., Darling, A.E. and Eisen, J.A., 2013. Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. *PloS one*, [online] 8(4), p.e62510. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0062510> [Accessed 16 Feb. 2016].

Lawley, T.D., Croucher, N.J., Yu, L., Clare, S., Sebaihia, M., Goulding, D., Pickard, D.J., Parkhill, J., Choudhary, J. and Dougan, G., 2009. Proteomic and genomic characterization of highly infectious *Clostridium difficile* 630 spores. *Journal of Bacteriology*, [online] 191(17), pp.5377–86. Available at: <http://jb.asm.org/content/191/17/5377.full> [Accessed 13 May 2015].

Lawson, P.A., Citron, D.M., Tyrrell, K.L. and Finegold, S.M., 2016. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe*, 40, pp.95–99.

Lazar, I., 2010. *Gel Analyzer*.

Legendre, P. and Lapointe, F., 2004. Assessing congruence among distance matrices: single-malt scotch whiskies revisited. *Australian & New Zealand Journal of Statistics*, 46(4), pp.615–629.

Leifert, C., Ball, K., Volakakis, N. and Cooper, J.M., 2008. Control of enteric pathogens in ready-to-eat vegetable crops in organic and 'low input'production systems: a HACCP-based approach. *Journal of Applied Microbiology*, 105(4), pp.931–950.

Letunic, I. and Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1), pp.W242–W245.

Leung, K. and Topp, E., 2001. Bacterial community dynamics in liquid swine manure during storage: molecular analysis using DGGE/PCR of 16S rDNA. *FEMS Microbiology Ecology*, [online] 38(2–3), pp.169–177. Available at: <http://doi.wiley.com/10.1111/j.1574-6941.2001.tb00895.x> [Accessed 17 Nov. 2014].

Lewis, C.J., 2007. Clostridial diseases. *Diseases of Sheep*, pp.156–167.

Lewis, C.J., 2011. Control of important clostridial diseases of sheep. *Veterinary Clinics: Food Animal Practice*, 27(1), pp.121–126.

Li, J., Sayeed, S. and McClane, B.A., 2007. Prevalence of enterotoxigenic *Clostridium perfringens* Isolates in Pittsburgh (Pennsylvania) area soils and home kitchens. *Applied and Environmental Microbiology*, [online] 73(22), pp.7218–24. Available at: <http://aem.asm.org/content/73/22/7218.short> [Accessed 7 Oct. 2014].

Liesack, W., Schnell, S. and Revsbech, N.P., 2000. Microbiology of flooded rice paddies. *FEMS Microbiology Reviews*, [online] 24(5), pp.625–645. Available at: <http://doi.wiley.com/10.1111/j.1574-6976.2000.tb00563.x> [Accessed 27 Oct. 2014].

Lindström, M., Heikinheimo, A., Lahti, P. and Korkeala, H., 2011. Novel insights into the epidemiology of *Clostridium perfringens* type A food poisoning. *Food Microbiology*, [online] 28(2), pp.192–8. Available at: <http://www.sciencedirect.com/science/article/pii/S0740002010000626> [Accessed 19 Sep. 2014].

Lindström, M., Keto, R., Markkula, A., Nevas, M., Hielm, S. and Korkeala, H., 2001. Multiplex PCR assay for detection and identification of *Clostridium botulinum* types A, B, E, and F in food and fecal material. *Applied and Environmental Microbiology*, [online] 67(12), pp.5694–9. Available at: <http://aem.asm.org/content/67/12/5694.short> [Accessed 10 Dec. 2014].

Lindström, M.K., Jankola, H.M., Hielm, S., Hyytiä, E.K. and Korkeala, H.J., 1999. Identification of *Clostridium botulinum* with API 20 A, Rapid ID 32 A and RapID ANA II. *FEMS Immunology & Medical Microbiology*, [online] 24(3), pp.267–274. Available at: <http://doi.wiley.com/10.1111/j.1574-695X.1999.tb01293.x> [Accessed 22 Jan. 2015].

Liu, X.-Z., Zhang, L.-M., Prosser, J.I. and He, J.-Z., 2009. Abundance and community structure of sulfate reducing prokaryotes in a paddy soil of southern China under different fertilization regimes. *Soil Biology and*

Biochemistry, [online] 41(4), pp.687–694. Available at: <http://www.sciencedirect.com/science/article/pii/S0038071709000108> [Accessed 27 Oct. 2014].

Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G. and Shao, Z., 2013. Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. *PloS one*, [online] 8(11), p.e80097. Available at: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0080097> [Accessed 3 Feb. 2016].

Ludwig, W., Schleifer, K.-H. and Whitman, W.B., 2009. Revised road map to the phylum Firmicutes. In: *Bergey's Manual® of Systematic Bacteriology*. Springer, pp.1–13.

Lúquez, C., Bianco, M.I., de Jong, L.I.T., Sagua, M.D., Arenas, G.N., Ciccarelli, A.S. and Fernández, R.A., 2005. Distribution of botulinum toxin-producing Clostridia in soils of Argentina. *Applied and Environmental Microbiology*, [online] 71(7), pp.4137–9. Available at: <http://aem.asm.org/content/71/7/4137> [Accessed 20 Oct. 2014].

Maddison, W. and Maddison, D., 2001. Mesquite: a modular system for evolutionary analysis.

del Mar Gamboa, M., Rodríguez, E. and Vargas, P., 2005. Diversity of mesophilic Clostridia in Costa Rican soils. *Anaerobe*, [online] 11(6), pp.322–6. Available at: <http://www.sciencedirect.com/science/article/pii/S1075996405000673> [Accessed 7 Oct. 2014].

Marsh, T.L., 1999. Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Current opinion in microbiology*, [online] 2(3), pp.323–7. Available at: <http://www.sciencedirect.com/science/article/pii/S1369527499800563> [Accessed 16 Oct. 2014].

Marzorati, M., Wittebolle, L., Boon, N., Daffonchio, D. and Verstraete, W., 2008. How to get more out of molecular fingerprints: practical tools for microbial ecology. *Environmental microbiology*, [online] 10(6), pp.1571–81. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18331337> [Accessed 13 Aug. 2014].

Matches, J.R., Liston, J. and Curran, D., 1974. *Clostridium perfringens* in the Environment. *Applied and Environmental Microbiology*, [online] 28(4), pp.655–660. Available at: <http://aem.asm.org/content/28/4/655.short> [Accessed 7 Oct. 2014].

Matsuda, K., Tsuji, H., Asahara, T., Kado, Y. and Nomoto, K., 2007. Sensitive quantitative detection of commensal bacteria by rRNA-targeted

reverse transcription-PCR. *Applied and Environmental Microbiology*, 73(1), pp.32–39.

Mawdsley, J.L., Bardgett, R.D., Merry, R.J., Pain, B.F. and Theodorou, M.K., 1995. Pathogens in livestock waste, their potential for movement through soil and environmental pollution. *Applied Soil Ecology*, [online] 2(1), pp.1–15. Available at: <http://www.sciencedirect.com/science/article/pii/092913939400039A> [Accessed 9 Oct. 2014].

McCafferty, J., Reid, R., Spencer, M., Hamp, T. and Fodor, A., 2012. P eak S tudio: a tool for the visualization and analysis of fragment analysis files. *Environmental Microbiology Reports*, 4(5), pp.556–561.

Mccarthy, H.E., French, N.P., Edwards, G.B., Miller, K. and Proudman, C.J., 2010. Why are certain premises at increased risk of equine grass sickness? A matched case-control study. *Equine Veterinary Journal*, [online] 36(2), pp.130–134. Available at: <http://doi.wiley.com/10.2746/0425164044868594> [Accessed 14 Oct. 2014].

McLauchlin, J. and Grant, K.A., 2007. *Clostridium botulinum* and Clostridium perfringens. In: *Foodborne Diseases*. Springer, pp.41–78.

De Medici, D., Anniballi, F., Wyatt, G.M., Lindström, M., Messelhäusser, U., Aldus, C.F., Delibato, E., Korkeala, H., Peck, M.W. and Fenicia, L., 2009. Multiplex PCR for detection of botulinum neurotoxin-producing clostridia in clinical, food, and environmental samples. *Applied and Environmental Microbiology*, [online] 75(20), pp.6457–61. Available at: <http://aem.asm.org/content/75/20/6457> [Accessed 28 Oct. 2014].

Meng, X., Yamakawa, K., Zou, K., Wang, X., Kuang, X., Lu, C., Wang, C., Karasawa, T. and Nakamura, S., 1999. Isolation and characterisation of neurotoxigenic *Clostridium butyricum* from soil in China. *Journal of Medical Microbiology*, [online] 48(2), pp.133–137. Available at: <http://jmm.sgmjournals.org/content/48/2/133.short> [Accessed 7 Oct. 2014].

Mignard, S. and Flandrois, J.P., 2006. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. *Journal of Microbiological Methods*, 67(3), pp.574–581.

Miller, M.A., Pfeiffer, W. and Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *2010 Gateway Computing Environments Workshop (GCE)*. [online] IEEE, pp.1–8. Available at: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5676129> [Accessed 7 Aug. 2015].

Minton, N.P. and Clarke, D.J., 1989. *Clostridia*. New York: Springer.

Miwa, T., 1975. Clostridia in soil of the Antarctica. *Japanese Journal of Medical Science & Biology*, [online] 28(4), pp.201–13. Available at: <http://europepmc.org/abstract/MED/175201> [Accessed 7 Oct. 2014].

Miyamoto, T., Kawahara, M. and Minamisawa, K., 2004. Novel endophytic nitrogen-fixing Clostridia from the grass *Miscanthus sinensis* as revealed by terminal restriction fragment length polymorphism analysis. *Applied and environmental microbiology*, [online] 70(11), pp.6580–6. Available at: <http://aem.asm.org/content/70/11/6580.short> [Accessed 6 Oct. 2014].

Moreno, L.I., Mills, D., Fetscher, J., John-Williams, K., Meadows-Jantz, L. and McCord, B., 2011. The application of amplicon length heterogeneity PCR (LH-PCR) for monitoring the dynamics of soil microbial communities associated with cadaver decomposition. *Journal of Microbiological Methods*, 84(3), pp.388–393.

Moriishi, K., Koura, M., Abe, N., Fujii, N., Fujinaga, Y., Inoue, K. and Ogumad, K., 1996. Mosaic structures of neurotoxins produced from *Clostridium botulinum* types C and D organisms. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, [online] 1307(2), pp.123–126. Available at: <http://www.sciencedirect.com/science/article/pii/0167478196000061> [Accessed 20 Aug. 2015].

Mosaddeghi, M.R., Mahboubi, A.A., Zandsalimi, S. and Unc, A., 2009. Influence of organic waste type and soil structure on the bacterial filtration rates in unsaturated intact soil columns. *Journal of Environmental Management*, [online] 90(2), pp.730–9. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=P1QOxsg5FdOtA18AXAz&page=1&doc=2> [Accessed 28 Oct. 2014].

Mowlick, S., Hirota, K., Takehara, T., Kaku, N., Ueki, K. and Ueki, A., 2012. Development of anaerobic bacterial community consisted of diverse clostridial species during biological soil disinfestation amended with plant biomass. *Soil science and plant nutrition*, 58(3), pp.273–287.

Moynihan, E.L., 2012. *Interactions between microbial community structure and pathogen survival in soil*. [online] Cranfield University. Available at: <https://dspace.lib.cranfield.ac.uk/handle/1826/7297> [Accessed 12 Nov. 2014].

Mukhopadhyay, A. and Mukhopadhyay, U.K., 2007. Novel multiplex PCR approaches for the simultaneous detection of human pathogens: Escherichia coli 0157: H7 and Listeria monocytogenes. *Journal of Microbiological Methods*, 68(1), pp.193–200.

Mulet, M., Lalucat, J. and García-Valdés, E., 2010. DNA sequence-based analysis of the *Pseudomonas* species. *Environmental Microbiology*, 12(6),

pp.1513–1530.

Munang, A., Hetron, M., Muyoyeta, P.M., Mweene, A.S. and Kida, H., 1996. Bovine clostridial infections in Zambia (1985-1994). *Japanese Journal of Veterinary Research*, 44(3), pp.175–178.

Nagashima, K., Hisada, T., Sato, M. and Mochizuki, J., 2003. Application of New Primer-Enzyme Combinations to Terminal Restriction Fragment Length Polymorphism Profiling of Bacterial Populations in Human Feces. *Applied and Environmental Microbiology*, [online] 69(2), pp.1251–1262. Available at: <http://aem.asm.org/content/69/2/1251.full> [Accessed 5 Jan. 2015].

Natsch, A., Keel, C., Troxler, J., Zala, M., VonAlbertini, N. and Defago, G., 1996. Importance of preferential flow and soil management in vertical transport of a biocontrol strain of *Pseudomonas fluorescens* in structured field soil. *Applied and Environmental Microbiology*, [online] 62(1), pp.33–40. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=P1QOxsg5FdOtA18AXAz&page=1&doc=9> [Accessed 28 Oct. 2014].

Neuhaus, J., Schrödl, W., Shehata, A.A. and Krüger, M., 2015. Detection of *Clostridium botulinum* in liquid manure and biogas plant wastes. *Folia Microbiologica*, [online] 60(5), pp.451–456. Available at: <http://link.springer.com/10.1007/s12223-015-0381-3> [Accessed 13 Apr. 2016].

Neuhaus, J., Shehata, A.A. and Krüger, M., 2015. Detection of pathogenic clostridia in biogas plant wastes. *Folia Microbiologica*, [online] 60(1), pp.15–19. Available at: <http://dx.doi.org/10.1007/s12223-014-0334-2>.

Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., Opsteegh, M., Langelaar, M., Threfall, J., Scheutz, F., van der Giessen, J. and Kruse, H., 2010. Food-borne diseases - the challenges of 20 years ago still persist while new ones continue to emerge. *International Journal of Food Microbiology*, [online] 139 Suppl, pp.S3-15. Available at: <http://www.sciencedirect.com/science/article/pii/S0168160510000383> [Accessed 11 Jul. 2014].

Newton, J.R., Hedderson, E.J., Adams, V.J., Mcgorum, B.C., Proudman, C.J. and Wood, J.L.N., 2010. An epidemiological study of risk factors associated with the recurrence of equine grass sickness (dysautonomia) on previously affected premises. *Equine Veterinary Journal*, [online] 36(2), pp.105–112. Available at: <http://doi.wiley.com/10.2746/0425164044868639> [Accessed 14 Oct. 2014].

Nguyen, N.-P., Warnow, T., Pop, M. and White, B., 2016. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms and Microbiomes*, 2, p.16004.

Nicholson, F.A., Groves, S.J. and Chambers, B.J., 2005. Pathogen survival during livestock manure storage and following land application. *Bioresource Technology*, [online] 96(2), pp.135–43. Available at: <http://www.sciencedirect.com/science/article/pii/S0960852404001087 > [Accessed 8 Oct. 2014].

Nilsson, W.B. and Strom, M.S., 2002. Detection and identification of bacterial pathogens of fish in kidney tissue using terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes. *Diseases of aquatic organisms*, 48(3), pp.175–185.

Noble, W., 1915. Experimental study of the distribution and habitat of the tetanus bacillus. *The Journal of Infectious Diseases*, pp.132–141.

Notermans, S., Dufrenne, J. and Oosterom, J., 1981. Persistence of *Clostridium botulinum* type B on a cattle farm after an outbreak of botulism. *Applied and Environmental Microbiology*, 41(1), pp.179–183.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C.M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V.S., Kodali, V.K., Li, W., Maglott, D., Masterson, P., McGarvey, K.M., Murphy, M.R., O'Neill, K., Pujar, S., Rangwala, S.H., Rausch, D., Riddick, L.D., Schoch, C., Shkeda, A., Storz, S.S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R.E., Vatsan, A.R., Wallin, C., Webb, D., Wu, W., Landrum, M.J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T.D. and Pruitt, K.D., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, [online] 44(D1), pp.D733-45. Available at: <http://nar.oxfordjournals.org/content/early/2015/11/06/nar.gkv1189.abstract> [Accessed 8 Mar. 2016].

Ogata, S. and Hongo, M., 1980. Bacteriophoges of the Genus Clostridium. *Advances in Applied Microbiology*, 25, p.241.

Ogle, D.H., 2015. *FSA: fisheries stock analysis*. *R package version 0.6*, .

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. and Wagner, H., 2013. Package 'vegan'. *Community Ecology Package*, 2(9).

de Oliveira Martins, L., Leal, E. and Kishino, H., 2008. Phylogenetic detection of recombination with a Bayesian prior on the distance between trees. *PLoS One*, 3(7), p.e2651.

Oliver, M.A. and Gregory, P.J., 2015. Soil, food security and human health: a review. *European Journal of Soil Science*, 66(2), pp.257–276.

Olsen, J.E. and Larsen, H.E., 1987. Bacterial decimation times in anaerobic digestions of animal slurries. *Biological Wastes*, [online] 21, pp.153–168. Available at: <http://www.sciencedirect.com/science/article/pii/0269748387901212>.

Östling, C.E. and Lindgren, S.E., 1991. Bacteria in manure and on manured and NPK-fertilised silage crops. *Journal of the Science of Food and Agriculture*, [online] 55(4), pp.579–588. Available at: <http://doi.wiley.com/10.1002/jsfa.2740550409> [Accessed 17 Nov. 2014].

Ouwerkerk, D. and Klieve, A.V., 2001. Bacterial Diversity within Feedlot Manure. *Anaerobe*, [online] 7(2), pp.59–66. Available at: <http://www.sciencedirect.com/science/article/pii/S1075996401903737> [Accessed 17 Nov. 2014].

Pachepsky, Y.A., Sadeghi, A.M., Bradford, S.A., Shelton, D.R., Guber, A.K. and Dao, T., 2006. Transport and fate of manure-borne pathogens: Modeling perspective. *Agricultural Water Management*, 86(1–2), pp.81–92.

Paradis, E., Claude, J. and Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), pp.289–290.

Parker, J., Fowler, N., Walmsley, M.L., Schmidt, T., Scharrer, J., Kowaleski, J., Grimes, T., Hoyos, S. and Chen, J., 2015. Analytical sensitivity comparison between singleplex real-time PCR and a multiplex PCR platform for detecting respiratory viruses. *PLoS One*, 10(11), p.e0143164.

Parte, A.C., 2018. LPSN-List of Prokaryotic names with Standing in Nomenclature (bacterio. net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology*, 68(6), pp.1825–1829.

Paulie, S., Perlmann, H. and Perlmann, P., 2006. Enzyme-linked Immunosorbent Assay. In: *Cell Biology*. Chichester, UK: Academic Press, pp.533–538.

Peck, M., Smith, T., Anniballi, F., Austin, J., Bano, L., Bradshaw, M., Cuervo, P., Cheng, L., Derman, Y. and Dorner, B., 2017. Historical perspectives and guidelines for botulinum neurotoxin subtype nomenclature. *Toxins*, 9(1), p.38.

Pei, A.Y., Oberdorf, W.E., Nossa, C.W., Agarwal, A., Chokshi, P., Gerz, E.A., Jin, Z., Lee, P., Yang, L., Poles, M., Brown, S.M., Sotero, S., DeSantis, T., Brodie, E., Nelson, K. and Pei, Z., 2010. Diversity of 16S

rRNA Genes within Individual Prokaryotic Genomes. *Applied and Environmental Microbiology*, [online] 76(12), pp.3886–3897. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2893482&tool=pmcentrez&rendertype=abstract> [Accessed 16 Feb. 2016].

Pepper, I.L., Gerba, C.P., Newby, D.T. and Rice, C.W., 2009. Soil: a public health threat or savior? *Critical Reviews in Environmental Science and Technology*, 39(5), pp.416–432.

Perry, D.A., 1982. Pectolytic *Clostridium* spp. in soils and rhizospheres of carrot and other arable crops in east Scotland. *Journal of Applied Bacteriology*, [online] 52(3), pp.403–408. Available at: <http://doi.wiley.com/10.1111/j.1365-2672.1982.tb05070.x> [Accessed 29 Oct. 2014].

Perry, D.A., 1985. Characteristics of pectolytic clostridia isolated from soil and plants in Scotland. *Journal of Applied Bacteriology*, [online] 58(3), pp.293–302. Available at: <http://doi.wiley.com/10.1111/j.1365-2672.1985.tb01463.x> [Accessed 29 Oct. 2014].

Persson, S., Torpdahl, M. and Olsen, K.E.P., 2008. New multiplex PCR method for the detection of *Clostridium difficile* toxin A (tcdA) and toxin B (tcdB) and the binary toxin (cdtA/cdtB) genes applied to a Danish strain collection. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, [online] 14(11), pp.1057–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19040478> [Accessed 8 Mar. 2015].

Peterson, L.R., Mehta, M.S., Patel, P.A., Hacek, D.M., Harazin, M., Nagwekar, P.P., Thomson, R.B. and Robicsek, A., 2011. Laboratory testing for *Clostridium difficile* infection: light at the end of the tunnel. *American Journal of Clinical Pathology*, [online] 136(3), pp.372–80. Available at: <http://ajcp.ascpjournals.org/content/136/3/372.short> [Accessed 19 Dec. 2014].

Petit, L., Gibert, M. and Popoff, M.R., 1999. *Clostridium perfringens*: toxinotype and genotype. *Trends in Microbiology*, [online] 7(3), pp.104–110. Available at: <http://www.sciencedirect.com/science/article/pii/S0966842X98014309> [Accessed 6 Oct. 2014].

Pett-Ridge, J. and Firestone, M.K., 2005. Redox fluctuation structures microbial communities in a wet tropical soil. *Applied and Environmental Microbiology*, [online] 71(11), pp.6998–7007. Available at: <http://aem.asm.org/cgi/doi/10.1128/AEM.71.11.6998-7007.2005> [Accessed 11 Apr. 2016].

Pichon, J.-P., Bonnaud, B., Cleuziat, P. and Mallet, F., 2006. Multiplex

degenerate PCR coupled with an oligo sorbent array for human endogenous retrovirus expression profiling. *Nucleic Acids Research*, 34(6), pp.e46–e46.

Pikuta, E. V, Itoh, T., Krader, P., Tang, J., Whitman, W.B. and Hoover, R.B., 2006. *Anaerovirgula multivorans* gen. nov., sp. nov., a novel spore-forming, alkaliphilic anaerobe isolated from Owens Lake, California, USA. *International Journal of Systematic and Evolutionary Microbiology*, [online] 56(Pt 11), pp.2623–9. Available at: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.64198-0> [Accessed 25 Feb. 2016].

Pirie, R.S., Jago, R.C. and Hudson, N.P.H., 2014. Equine grass sickness. *Equine Veterinary Journal*, [online] 46(5), pp.545–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24580639> [Accessed 8 Apr. 2015].

Poehlein, A., Cebulla, M., Ilg, M.M., Bengelsdorf, F.R., Schiel-Bengelsdorf, B., Whited, G., Andreesen, J.R., Gottschalk, G., Daniel, R. and Dürre, P., 2015. The Complete Genome Sequence of *Clostridium aceticum*: a Missing Link between Rnf- and Cytochrome-Containing Autotrophic Acetogens. *mBio*, [online] 6(5), pp.e01168-15. Available at: <http://mbio.asm.org/content/6/5/e01168-15.full> [Accessed 24 Feb. 2016].

Popoff, M.R. and Bouvet, P., 2013. Genetic characteristics of toxigenic Clostridia and toxin gene evolution. *Toxicon*, [online] 75, pp.63–89. Available at: <https://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=Y2kneCfSsV2zMNQjBkS&page=1&doc=10> [Accessed 28 Sep. 2014].

Poritz, M.A., Blaschke, A.J., Byington, C.L., Meyers, L., Nilsson, K., Jones, D.E., Thatcher, S.A., Robbins, T., Lingenfelter, B. and Amiott, E., 2011. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection. *PLoS One*, 6(10), p.e26047.

Potts, J.M. and Elith, J., 2006. Comparing species abundance models. *Ecological Modelling*, 199(2), pp.153–163.

Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G. and Toth, I.K., 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods*, 8(1), pp.12–24.

Pulvirenti, A., Ronga, D., Zaghi, M., Tomasselli, A.R., Mannella, L. and Pecchioni, N., 2015. Pelleting is a successful method to eliminate the presence of *Clostridium spp.* from the digestate of biogas plants. *Biomass and Bioenergy*, 76(14) pp298-299

Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., Yang, Y. and Zhang, C., 2012. MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. *Nucleic Acids Research*, [online] 40(Web Server issue), pp.W205-8. Available at: <http://nar.oxfordjournals.org/content/early/2012/06/09/nar.gks552.full> [Accessed 6 Apr. 2016].

Ramirez, N. and Abel-Santos, E., 2010. Requirements for germination of *Clostridium sordellii* spores in vitro. *Journal of Bacteriology*, [online] 192(2), pp.418–25. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2805323&tool=pmcentrez&rendertype=abstract> [Accessed 11 Dec. 2014].

Reddy, K.R., Khaleel, R. and Overcash, M.R., 1981. Behavior and Transport of Microbial Pathogens and Indicator Organisms in Soils Treated with Organic Wastes 1. *Journal of Environmental Quality*, 10(3), pp.255–266.

Reeves, P.R., Farnell, L. and Lan, R., 1994. MULTICOMP: a program for preparing sequence data for phylogenetic analysis. *Bioinformatics*, 10(3), pp.281–284.

Rice, W.A. and Paul, E.A., 1972. The organisms and biological processes involved in asymbiotic nitrogen fixation in waterlogged soil amended with straw. *Canadian Journal of Microbiology*, [online] 18(6), pp.715–723. Available at: <http://www.nrcresearchpress.com/doi/abs/10.1139/m72-114>.

Robinson, D.F. and Foulds, L.R., 1981. Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1–2), pp.131–147.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G. and Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal*, [online] 1(4), pp.283–90. Available at: <http://dx.doi.org/10.1038/ismej.2007.53> [Accessed 9 Jul. 2014].

Rogers, S.W., Donnelly, M., Peed, L., Kelty, C.A., Mondal, S., Zhong, Z. and Shanks, O.C., 2011. Decay of bacterial pathogens, fecal indicators, and real-time quantitative PCR genetic markers in manure-amended soils. *Applied and Environmental Microbiology*, 77(14), pp.4839–4848.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. and Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, [online] 61(3), pp.539–42. Available at: <http://sysbio.oxfordjournals.org/content/61/3/539.short> [Accessed 10 Jul. 2014].

Rood, J.I. and Cole, S.T., 1991. Molecular genetics and pathogenesis of *Clostridium perfringens. Microbiology and Molecular Biology Reviews*, [online] 55(4), pp.621–648. Available at: <http://mmbr.asm.org/content/55/4/621.short> [Accessed 6 Oct. 2014].

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. and Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19(18), pp.2496–2497.

Ruengwilysup, C., Detvisitsakun, C., Aumyat, N. and Fung, D.Y.C., 2009. Application of a colony pcr technique with fung's double tube method for rapid detection and confirmation of *Clostridium perfringens*. *Journal of Rapid Methods & Automation in Microbiology*, 17(3), pp.280–290.

Russo, S.E., Legge, R., Weber, K.A., Brodie, E.L., Goldfarb, K.C., Benson, A.K. and Tan, S., 2012. Bacterial community structure of contrasting soils underlying Bornean rain forests: Inferences from microarray and next-generation sequencing methods. *Soil Biology and Biochemistry*, [online] 55, pp.48–59. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=21&SID=R19cWLD57gqoWihBjlx&page=1&doc=42> [Accessed 15 Oct. 2014].

Safadoust, A., Mahboubi, A.A., Gharabaghi, B., Mosaddeghi, M.R., Voroney, P., Unc, A. and Sayyad, G., 2011. Bacterial filtration rates in repacked and weathered soil columns. *Geoderma*, [online] 167–168, pp.204–213. Available at: <http://www.sciencedirect.com/science/article/pii/S0016706111002576> [Accessed 28 Oct. 2014].

Sahlström, L., Bagge, E., Emmoth, E., Holmqvist, A., Danielsson-Tham, M.-L. and Albihn, A., 2008. A laboratory study of survival of selected microorganisms after heat treatment of biowaste used in biogas plants. *Bioresource Technology*, [online] 99, pp.7859–7865. Available at: <http://www.sciencedirect.com/science/article/pii/S0960852407008164>.

Al Saif, N. and Brazier, J.S., 1996. The distribution of *Clostridium difficile* in the environment of South Wales. *Journal of Medical Microbiology*, [online] 45(2), pp.133–137. Available at: <http://jmm.sgmjournals.org/content/45/2/133.short> [Accessed 7 Oct. 2014].

Salipante, S.J., SenGupta, D.J., Cummings, L.A., Land, T.A., Hoogestraat, D.R. and Cookson, B.T., 2015. Application of whole genome sequencing for bacterial strain typing in molecular epidemiology. *Journal of Clinical Microbiology*, p.JCM-03385.

Saltykova, A., Wuyts, V., Mattheus, W., Bertrand, S., Roosens, N.H.C.,

Marchal, K. and De Keersmaecker, S.C.J., 2018. Comparison of SNP-based subtyping workflows for bacterial isolates using WGS data, applied to Salmonella enterica serotype Typhimurium and serotype 1, 4,[5], 12: i:-. *PLoS One*, 13(2), p.e0192504.

Sandler, R.J., Rocke, T.E., Samuel, M.D. and Yuill, T.M., 1993. Seasonal prevalence of *Clostridium botulinum* type C in sediments of a northern California wetland. *Journal of Wildlife Diseases*, 29(4), pp.533–539.

Sandler, R.J., Rocke, T.E. and Yuill, T.M., 1998. The inhibition of *Clostridium botulinum* type C by other bacteria in wetland sediments. *Journal of Wildlife Diseases*, 34(4), pp.830–833.

Sasaki, Y., Kojima, A., Aoki, H., Ogikubo, Y., Takikawa, N. and Tamura, Y., 2002. Phylogenetic analysis and PCR detection of *Clostridium chauvoei*, *Clostridium haemolyticum, Clostridium novyi* types A and B, and *Clostridium septicum* based on the flagellin gene. *Veterinary Microbiology*, [online] 86(3), pp.257–267. Available at: <http://www.sciencedirect.com/science/article/pii/S0378113502000020> [Accessed 15 Jan. 2015].

Sasaki, Y., Yamamoto, K., Kojima, A., Norimatsu, M. and Tamura, Y., 2000a. Rapid identification and differentiation of pathogenic clostridia in gas gangrene by polymerase chain reaction based on the 16S-23S rDNA spacer region. *Research in veterinary science*, [online] 69(3), pp.289–94. Available at: <http://www.sciencedirect.com/science/article/pii/S0034528800904314> [Accessed 15 Jan. 2015].

Sasaki, Y., Yamamoto, K., Kojima, A., Tetsuka, Y., Norimatsu, M. and Tamura, Y., 2000b. Rapid and Direct Detection of *Clostridium chauvoei* by PCR of the 16S-23S rDNA Spacer Region and Partial 23S rDNA Sequences. *Journal of Veterinary Medical Science*, [online] 62(12), pp.1275–1281. Available at: <http://europepmc.org/abstract/MED/11193343> [Accessed 15 Jan. 2015].

Sathish, S. and Swaminathan, K., 2009. Genetic diversity among toxigenic clostridia isolated from soil, water, meat and associated polluted sites in South India. *Indian Journal of Medical Microbiology*, 27(4), p.311.

Sawada, H., Suzuki, F., Matsuda, I. and Saitou, N., 1999. Phylogenetic Analysis of *Pseudomonas syringae* Pathovars Suggests the Horizontal Gene Transfer of argK and the Evolutionary Stability of hrp Gene Cluster. *Journal of Molecular Evolution*, [online] 49(5), pp.627–644. Available at: <http://link.springer.com/10.1007/PL00006584> [Accessed 9 Feb. 2016].

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*,

27(4), p.592.

Schnurer, A., Schink, B. and Svensson, B.H., 1996. *Clostridium ultunense* sp. nov., a mesophilic bacterium oxidizing acetate in syntrophic association with a hydrogenotrophic methanogenic bacterium. *International Journal of Systematic Bacteriology*, [online] 46(4), pp.1145–52. Available at: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-46-4-1145> [Accessed 2 Mar. 2016].

Schoenholz, P. and Meyer, K.F., 1922. The occurrence of the spores of *B. botulinus* in the Hawaiian Islands and China. VII. *The Journal of Infectious Diseases*, pp.610–613.

Schütte, U.M.E., Abdo, Z., Bent, S.J., Shyu, C., Williams, C.J., Pierson, J.D. and Forney, L.J., 2008. Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Applied Microbiology and Biotechnology*, 80(3), pp.365–380.

Scottish, G., 2017. *Farm Livestock - All Livestock*. [online] Available at: <https://www2.gov.scot/Topics/Statistics/Browse/Agriculture-Fisheries/agritopics/alllivestock> [Accessed 4 Mar. 2019].

Sebaihia, M., Peck, M.W., Minton, N.P., Thomson, N.R., Holden, M.T.G., Mitchell, W.J., Carter, A.T., Bentley, S.D., Mason, D.R., Crossman, L., Paul, C.J., Ivens, A., Wells-Bennik, M.H.J., Davis, I.J., Cerdeño-Tárraga, A.M., Churcher, C., Quail, M.A., Chillingworth, T., Feltwell, T., Fraser, A., Goodhead, I., Hance, Z., Jagels, K., Larke, N., Maddison, M., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., White, B., Whithead, S. and Parkhill, J., 2007. Genome sequence of a proteolytic (Group I) C*lostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Research*, [online] 17(7), pp.1082–92. Available at: <http://genome.cshlp.org/content/17/7/1082.short> [Accessed 7 Apr. 2015].

Sedlackova, V., Dziedzinska, R., Babak, V. and Kralik, P., 2017. The detection and quantification of *Bacillus thuringiensis* spores from soil and swabs using quantitative PCR as a model system for routine diagnostics of Bacillus anthracis. *Journal of Applied Microbiology*, 123(1), pp.116–123.

Sedlar, K., Kolek, J., Skutkova, H., Branska, B., Provaznik, I. and Patakova, P., 2015. Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol. *Journal of Biotechnology*, [online] 214, pp.113–4. Available at: <http://www.sciencedirect.com/science/article/pii/S0168165615301279> [Accessed 8 Jan. 2016].

Sedlar, K., Škutková, H., Kolek, J., Patáková, P. and Provaznik, I., 2014. Identification and characteriazation of sol operon in *Clostridium pasteurianum* NRRL B-598 genome. In: *2nd International Conference on Chemical Technology ICCT2014*. [online] Available at: <https://www.researchgate.net/publication/265058321_Identification_and_characteriazation_of_sol_operon_in_Clostridium_pasteurianum_NRRL_B-598_genome> [Accessed 26 Feb. 2016].

Serikawa, T., Nakamura, S. and Nishida, S., 1977. Distribution of Clostridium botulinum Type C in Ishikawa Prefecture, and Applicability of Agglutination to Identification of Nontoxigenic Isolates of *C. botulinum* Type C. *Microbiology and Immunology*, [online] 21(3), pp.127–136. Available at: <http://doi.wiley.com/10.1111/j.1348-0421.1977.tb00274.x> [Accessed 20 Oct. 2014].

Setlow, P., 2007. I will survive: DNA protection in bacterial spores. *Trends in Microbiology*, [online] 15(4), pp.172–80. Available at: <http://www.sciencedirect.com/science/article/pii/S0966842X07000261> [Accessed 23 Nov. 2014].

Shehata, A., Schrödl, W., Neuhaus, J. and Krüger, M., 2013. Antagonistic effect of different bacteria on *Clostridium botulinum* types A, B, D and E in vitro. *The Veterinary Record*, [online] 172(2), p.47. Available at: <http://veterinaryrecord.bmj.com/content/early/2012/12/17/vr.101184.full> [Accessed 19 Dec. 2014].

Simango, C., 2006. Prevalence of *Clostridium difficile* in the environment in a rural community in Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, [online] 100(12), pp.1146–50. Available at: <http://www.sciencedirect.com/science/article/pii/S0035920306000976> [Accessed 19 Jan. 2015].

Skarin, H. and Segerman, B., 2014. Plasmidome Interchange between *Clostridium botulinum, Clostridium novyi* and Clostridium *haemolyticum* Converts Strains of Independent Lineages into Distinctly Different Pathogens. *PloS one*, [online] 9(9), p.e107777. Available at: <http://dx.doi.org/10.1371%2Fjournal.pone.0107777>.

Smalla, K., Oros-Sichler, M., Milling, A., Heuer, H., Baumgarte, S., Becker, R., Neuber, G., Kropf, S., Ulrich, A. and Tebbe, C.C., 2007. Bacterial diversity of soils assessed by DGGE, T-RFLP and SSCP fingerprints of PCR-amplified 16S rRNA gene fragments: do the different methods provide similar results? *Journal of microbiological methods*, [online] 69(3), pp.470–9. Available at: <http://www.sciencedirect.com/science/article/pii/S0167701207000681> [Accessed 19 Sep. 2014].

Smit, E., Leeflang, P., Gommans, S., van den Broek, J., van Mil, S. and

Wernars, K., 2001. Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Applied and Environmental Microbiology*, 67(5), pp.2284–2291.

Smith, G. and Milligan, R., 1979. *Clostridium botulinum* in soil on the site of the former Metropolitan (Caledonian) Cattle Market, London. *Journal of Hygiene*, [online] 83(02), p.237. Available at: <http://journals.cambridge.org/abstract_S0022172400026024> [Accessed 20 Oct. 2014].

Smith, G., Milligan, R. and Moryson, C.J., 1978. *Clostridium botulinum* in aquatic environments in Great Britain and Ireland. *Journal of Hygiene*, [online] 80(03), pp.431–438. Available at: <http://journals.cambridge.org/abstract_S0022172400024906> [Accessed 27 Oct. 2014].

Smith, G. and Young, A., 1980. *Clostridium botulinum* in British soil. *Journal of Hygiene*, [online] 85(02), p.271. Available at: <http://journals.cambridge.org/abstract_S0022172400063300> [Accessed 20 Oct. 2014].

Smith, L.D., 1975a. Common Mesophilic Anaerobes, Including *Clostridium botulinum* and *Clostridium tetani*, in 21 Soil Specimens. *Applied and Environmental Microbiology*, [online] 29(5), pp.590–594. Available at: <http://aem.asm.org/content/29/5/590.abstract> [Accessed 21 Oct. 2014].

Smith, L.D., 1975b. Inhibition of *Clostridium botulinum* by Strains of *Clostridium perfringens* Isolated from Soil. *Applied and Environmental Microbiology*, [online] 30(2), pp.319–323. Available at: <http://aem.asm.org/content/30/2/319.short> [Accessed 7 Oct. 2014].

Smith, L.D., 1978. The occurrence of *Clostridium botulinum* and *Clostridium tetani* in the soil of the United States. *Health Laboratory Science*, [online] 15(2), pp.74–80. Available at: <http://europepmc.org/abstract/MED/355208> [Accessed 20 Oct. 2014].

Smith, T.J., Hill, K.K., Xie, G., Foley, B.T., Williamson, C.H.D., Foster, J.T., Johnson, S.L., Chertkov, O., Teshima, H., Gibbons, H.S., Johnsky, L.A., Karavis, M.A. and Smith, L.A., 2015. Genomic sequences of six botulinum neurotoxin-producing strains representing three clostridial species illustrate the mobility and diversity of botulinum neurotoxin genes. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, [online] 30, pp.102–13. Available at: <http://www.sciencedirect.com/science/article/pii/S1567134814004481> [Accessed 18 Sep. 2015].

Sobhy, H. and Colson, P., 2012. Gemi: PCR primers prediction from

multiple alignments. *Comparative and Functional Genomics*, [online] 2012, p.783138. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3535827&tool=pmcentrez&rendertype=abstract> [Accessed 12 Jun. 2015].

Soler, L., Yáñez, M.A., Chacon, M.R., Aguilera-Arreola, M.G., Catalán, V., Figueras, M.J. and Martínez-Murcia, A.J., 2004. Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. *International Journal of Systematic and Evolutionary Microbiology*, [online] 54(Pt 5), pp.1511–9. Available at: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.03048-0> [Accessed 19 Jan. 2016].

Song, Y.-L., Kato, N., Liu, C.-X., Matsumiya, Y., Kato, H. and Watanabe, K., 2000. Rapid identification of 11 human intestinal *Lactobacillus* species by multiplex PCR assays using group- and species-specific primers derived from the 16S–23S rRNA intergenic spacer region and its flanking 23S rRNA. *FEMS Microbiology Letters*, [online] 187(2), pp.167–173. Available at: <http://dx.doi.org/10.1111/j.1574-6968.2000.tb09155.x>.

Song, Y., Liu, C. and Finegold, S.M., 2004. Real-time PCR quantitation of Clostridia in feces of autistic children. *Applied and Environmental Microbiology*, [online] 70(11), pp.6459–65. Available at: <http://aem.asm.org/content/70/11/6459.short> [Accessed 6 Oct. 2014].

Song, Y., Liu, C., Molitoris, D., Tomzynski, T.J., Mc Teague, M., Read, E. and Finegold, S.M., 2002. Use of 16S–23S rRNA spacer-region (SR)-PCR for identification of intestinal Clostridia. *Systematic and Applied Microbiology*, 25(4), pp.528–535.

Songer, J.G., 1996. Clostridial enteric diseases of domestic animals. *Clinical Microbiology Reviews*, [online] 9(2), pp.216–34. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=172891&tool=pmcentrez&rendertype=abstract> [Accessed 19 Jan. 2015].

Songer, J.G., 2010. Clostridia as agents of zoonotic disease. *Veterinary Microbiology*, [online] 140(3–4), pp.399–404. Available at: <http://www.sciencedirect.com/science/article/pii/S0378113509003241> [Accessed 19 Sep. 2014].

Sonnabend, W.F., Sonnabend, U.P. and Krech, T., 1987. Isolation of *Clostridium botulinum* type G from Swiss soil specimens by using sequential steps in an identification scheme. *Applied and Environmental Microbiology*, [online] 53(8), pp.1880–1884. Available at: <http://aem.asm.org/content/53/8/1880.short> [Accessed 20 Oct. 2014].

South, V., 2014. Clostridial diseases of the horse. *In Practice*, [online] 36(1), pp.27–33. Available at:

<http://inpractice.bmj.com/content/36/1/27> [Accessed 19 Dec. 2014].

Stackebrandt, E., Kramer, I., Swiderski, J. and Hippe, H., 1999. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium. FEMS Immunology & Medical Microbiology*, [online] 24(3), pp.253–258. Available at: <http://doi.wiley.com/10.1111/j.1574-695X.1999.tb01291.x> [Accessed 26 Nov. 2014].

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, [online] 30(9), pp.1312–3. Available at: <http://bioinformatics.oxfordjournals.org/content/early/2014/02/07/bioinformatics.btu033.full> [Accessed 15 Jul. 2014].

Stamatakis, A., Hoover, P. and Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, [online] 57(5), pp.758–771. Available at: <http://sysbio.oxfordjournals.org/content/57/5/758.short> [Accessed 30 Jun. 2015].

Steel, M.A. and Penny, D., 1993. Distributions of tree comparison metrics—some new results. *Systematic Biology*, 42(2), pp.126–141.

Stefanis, C., Alexopoulos, A., Voidarou, C., Vavias, S. and Bezirtzoglou, E., 2014. Prevalence of *C. perfringens* in agricultural fields with different vegetation types. *Folia Microbiologica*, [online] 59(1), pp.1–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23686361> [Accessed 20 Oct. 2014].

Stoddard, C.S., Coyne, M.S. and Grove, J.H., 1998. Fecal Bacteria Survival and Infiltration through a Shallow Agricultural Soil: Timing and Tillage Effects. *Journal of Environment Quality*, [online] 27(6), p.1516. Available at: <https://dl.sciencesocieties.org/publications/jeq/abstracts/27/6/JEQ0270061516> [Accessed 10 Nov. 2014].

Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28(6), pp.1102–1104.

Stringer, S.C., Carter, A.T., Webb, M.D., Wachnicka, E., Crossman, L.C., Sebaihia, M. and Peck, M.W., 2013. Genomic and physiological variability within Group II (non-proteolytic) *Clostridium botulinum. BMC Genomics*, [online] 14(1), p.333. Available at: <http://www.biomedcentral.com/1471-2164/14/333> [Accessed 10 Nov. 2014].

Stubbs, S.L.J., Brazier, J.S., O'Neill, G.L. and Duerden, B.I., 1999. PCR Targeted to the 16S-23S rRNA Gene Intergenic Spacer Region of *Clostridium difficile* and Construction of a Library Consisting of 116

Different PCR Ribotypes. *Journal of Clinical Microbiology*, 37(2), pp.461–463.

Swofford, D.L., 1998. *Paup 4.0 beta version for windows: phylogenetic analysis using parsimony*.

Tabaqchali, S. and Jumaa, P., 1995. Diagnosis and management of *Clostridium difficile* infection. *BMJ (Clinical research ed.)*, [online] 310(6991), pp.1375–80. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2549751&tool=pmcentrez&rendertype=abstract> [Accessed 20 Oct. 2014].

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), pp.585–595.

Tambong, J.T., Xu, R., Kaneza, C.-A. and Nshogozabahizi, J.-C., 2014. An In-depth analysis of a multilocus phylogeny identifies LeuS as a reliable phylogenetic marker for the genus *Pantoea*. *Evolutionary Bioinformatics*, [online] 10, p.EBO.S15738. Available at: <https://doi.org/10.4137/EBO.S15738>.

Tanner, R.S., Miller, L.M. and Yang, D., 1993. *Clostridium ljungdahlii s*p. nov., an acetogenic species in clostridial rRNA homology group I. *International Journal of Systematic Bacteriology*, [online] 43(2), pp.232–6. Available at: <http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-43-2-232> [Accessed 24 Feb. 2016].

Tazzyman, S.J. and Bonhoeffer, S., 2014. Why there are no essential genes on plasmids. *Molecular Biology and Evolution*, p.msu293.

Thurston-Enriquez, J., Gilley, J. and Eghball, B., 2005. Microbial quality of runoff following land application of cattle manure and swine slurry. *Journal of Water and Health*, [online] 3(2), pp.157–171. Available at: <http://www.iwaponline.com/jwh/003/jwh0030157.htm> [Accessed 10 Nov. 2014].

Thwaites, C.L., Beeching, N.J. and Newton, C.R., 2015. Maternal and neonatal tetanus. *The Lancet*, 385(9965), pp.362–370.

Tiedje, J.M., Asuming-Brempong, S., Nüsslein, K., Marsh, T.L. and Flynn, S.J., 1999. Opening the black box of soil microbial diversity. *Applied Soil Ecology*, [online] 13(2), pp.109–122. Available at: <http://www.sciencedirect.com/science/article/pii/S0929139399000268> [Accessed 27 Jan. 2015].

Tiedje, J.M., Sexstone, A.J., Parkin, T.B. and Revsbech, N.P., 1984. Anaerobic processes in soil. *Plant and Soil*, 76(1–3), pp.197–212.

Tillotson, K., Traub-Dargatz, J.L., Dickinson, C.E., Ellis, R.P., Morley, P.S.,

Hyatt, D.R., Magnuson, R.J., Riddle, W.T., Bolte, D. and Salman, M.D., 2002. Population-based study of fecal shedding of *Clostridium perfringens* in broodmares and foals. *Journal of the American Veterinary Medical Association*, [online] 220(3), pp.342–348. Available at: <http://avmajournals.avma.org/doi/abs/10.2460/javma.2002.220.342> [Accessed 19 Jan. 2015].

Torniainen, M., Sahlström, L., Maunuksela, L. and Paavola, T., 2011. Survival of pathogenic bacteria in different anaerobic treatment processes. In: *In Animal hygiene and sustainable livestock production. Proceedings of the XVth International Congress of the International Society for Animal Hygiene*. [online] Vienna: Tribun EU, pp.1103–1104. Available at: <http://www.cabdirect.org/abstracts/20113298491.html>.

Trunet, C., Carlin, F. and Coroller, L., 2017. Investigating germination and outgrowth of bacterial spores at several scales. *Trends in Food Science & Technology*, 64, pp.60–68.

Ulrich, A. and Becker, R., 2006. Soil parent material is a key determinant of the bacterial community structure in arable soils. *FEMS Microbiology Ecology*, [online] 56(3), pp.430–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16689875> [Accessed 16 Oct. 2014].

Unc, A. and Goss, M.J., 2004. Transport of bacteria from manure and protection of water resources. *Applied Soil Ecology*, 25(1), pp.1–18.

Useh, N.M., Ibrahim, N.D.G., Nok, A.J. and Esievo, K.A.N., 2006. Relationship between outbreaks of blackleg in cattle and annual rainfall in Zaria, Nigeria. *The Veterinary Record*, 158(3), p.100.

Useh, N.M., Nok, A.J. and Esievo, K.A.N., 2006. Blackleg in ruminants. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 1, pp.1–8.

Vaidya, G., Lohman, D.J. and Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, [online] 27(2), pp.171–180. Available at: <http://doi.wiley.com/10.1111/j.1096-0031.2010.00329.x> [Accessed 30 Jun. 2016].

Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K., Kyrpides, N.C. and Pati, A., 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Research*, 43(14), pp.6761–6771.

Venables, W.N. and Ripley, B.D., 2002. Tree-based methods. In: *Modern Applied Statistics with S-Plus*. Springer, pp.251–269.

Venglovsky, J., Martinez, J. and Placha, I., 2006. Hygienic and ecological

risks connected with utilization of animal manures and biosolids in agriculture. *Livestock Science*, [online] 102(3), pp.197–203. Available at: <http://www.sciencedirect.com/science/article/pii/S1871141306000990 > [Accessed 9 Nov. 2014].

Vidal, D., Anza, I., Taggart, M.A., Pérez-Ramírez, E., Crespo, E., Hofle, U. and Mateo, R., 2013. Environmental factors influencing the prevalence of a *Clostridium botulinum* type C/D mosaic strain in nonpermanent Mediterranean wetlands. *Applied and Environmental Microbiology*, [online] 79(14), pp.4264–71. Available at: <http://aem.asm.org/content/79/14/4264.short> [Accessed 10 Dec. 2014].

Vijayavel, K. and Kashian, D.R., 2014. Evaluation of *Clostridium perfringens* as a tracer of sewage contamination in sediments by two enumeration methods. *Environmental Monitoring and Assessment*, [online] 186(9), pp.5617–24. Available at: <https://apps.webofknowledge.com/full_record.do?product=UA&search_ mode=GeneralSearch&qid=1&SID=Q16fbmnKbJ4NDjhkPFR&page=1&doc =1> [Accessed 1 Oct. 2014].

Voidarou, C., Bezirtzoglou, E., Alexopoulos, A., Plessas, S., Stefanis, C., Papadopoulos, I., Vavias, S., Stavropoulou, E., Fotou, K., Tzora, A. and Skoufos, I., 2011. Occurrence of *Clostridium perfringens* from different cultivated soils. *Anaerobe*, [online] 17(6), pp.320–4. Available at: <http://www.sciencedirect.com/science/article/pii/S1075996411000862 > [Accessed 30 Oct. 2014].

Waggett, B.E., McGorum, B.C., Wernery, U., Shaw, D.J. and Pirie, R.S., 2010. Prevalence of *Clostridium perfringens* in faeces and ileal contents from grass sickness affected horses: comparisons with 3 control populations. *Equine Veterinary Journal*, [online] 42(6), pp.494–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20716188> [Accessed 19 Jan. 2015].

Ward, B. and Hicks, N., 2013. *Recent and future changes in the global and UK climate*. [online] Available at: <http://www.lse.ac.uk/GranthamInstitute/wp-content/uploads/2014/03/changes-in-global-and-uk-climate-low-res.pdf>.

Weber, S., Stubner, S. and Conrad, R., 2001. Bacterial populations colonizing and degrading rice straw in anoxic paddy soil. *Applied and Environmental Microbiology*, [online] 67(3), pp.1318–27. Available at: <http://aem.asm.org/content/67/3/1318.short> [Accessed 27 Oct. 2014].

Weigand, M.R., Pena-Gonzalez, A., Shirey, T.B., Broeker, R.G., Ishaq, M.K., Konstantinidis, K.T. and Raphael, B.H., 2015. Implications of

Genome-Based Discrimination between *Clostridium botulinum* Group I and *Clostridium sporogenes* Strains for Bacterial Taxonomy. *Applied and Environmental Microbiology*, [online] 81(16), pp.5420–9. Available at: <http://aem.asm.org/content/81/16/5420.full> [Accessed 25 Feb. 2016].

Weigel, J., Tanner, R. and Rainey, F., 2006. An introduction to the family Clostridiaceae. In: M. Dworkin, S. Falkow, E. Rosenberg, K.H. Schleifer and E. Stackebrandt, eds., *The Prokaryotes: a Handbook on the Biology of Bacteria*, 3rd ed. Springer New York, pp.654–678.

Wheeldon, L.J., Worthington, T., Hilton, A.C., Elliott, T.S.J. and Lambert, P.A., 2008. Physical and chemical factors influencing the germination of *Clostridium difficile* spores. *Journal of Applied Microbiology*, [online] 105(6), pp.2223–30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19120667> [Accessed 11 Dec. 2014].

Whitman, W.B. and Parte, A.C., 2009. *Systematic bacteriology*. Springer.

WHO, 2016. *WHO vaccine-preventable diseases: monitoring system 2016 global summary*. [online] Available at: <http://apps.who.int/immunization_monitoring/globalsummary/timeseries/tsincidencettetanus.html>.

Wickham, H., 2007. Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12), pp.1–20.

Wickham, H., 2016. *ggplot2: elegant graphics for data analysis*. Springer.

Wickham, H. and Wickham, M.H., 2018. *Package 'plyr'*.

Wiencek, K.M., Klapes, N.A. and Foegeding, P.M., 1990. Hydrophobicity of *Bacillus* and *Clostridium* spores. *Applied and Environmental Microbiology*, 56(9), pp.2600–2605.

Wilkins, C., Richter, M., Hobbs, W., Whitcomb, M., Bergh, N. and Carstens, J., 1988. Occurrence of *Clostridium* in soil and horses. *South Arfican Medical Journal*, [online] 73(12), pp.718–720. Available at: <http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=19&SID=Z2JK7yiyr85WEdtgqD3&page=1&doc=19> [Accessed 21 Oct. 2014].

Williamson, J.L., Rocke, T.E. and Aiken, J.M., 1999. In situ detection of the *Clostridium botulinum* type C1 toxin gene in wetland sediments with a nested PCR assay. *Applied and Environmental Microbiology*, [online] 65(7), pp.3240–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10388729> [Accessed 11 Apr. 2016].

Wintzingerode, F. V, Göbel, U.B. and Stackebrandt, E., 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, [online] 21(3), pp.213–229. Available at: <http://doi.wiley.com/10.1111/j.1574-6976.1997.tb00351.x> [Accessed 16 Oct. 2014].

Wolf, R., Hiesel, J., Kuchling, S., Deutz, A., Kastelic, J., Barkema, H.W. and Wagner, P., 2017. Spatial-temporal cluster analysis of fatal *Clostridium chauvoei* cases among cattle in Styria, Austria between 1986 and 2013. *Preventive Veterinary Medicine*, 138, pp.134–138.

Wood, J.L.N., Milne, E.M. and Doxey, D.L., 1998. A case-control study of grass sickness (equine dysautonomia) in the United Kingdom. *The Veterinary Journal*, [online] 156(1), pp.7–14. Available at: <http://www.sciencedirect.com/science/article/pii/S1090023398800555>.

Wylie, C.E. and Proudman, C.J., 2009. Equine grass sickness: epidemiology, diagnosis, and global distribution. *The Veterinary Clinics of North America. Equine practice*, [online] 25(2), pp.381–99. Available at: <http://www.sciencedirect.com/science/article/pii/S074907390900039X> [Accessed 14 Oct. 2014].

Wylie, C.E., Shaw, D.J., Fordyce, F.M., Lilly, A. and McGorum, B.C., 2014. Equine grass sickness in Scotland: a case-control study of signalment- and meteorology-related risk factors. *Equine Veterinary Journal*, [online] 46(1), pp.64–71. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23662745> [Accessed 27 Jul. 2015].

Xu, C., Wang, D., Huber, A., Weese, S.J. and Warriner, K., 2016. Persistence of *Clostridium difficile* in wastewater treatment-derived biosolids during land application or windrow composting. *Journal of Applied Microbiology*, [online] 120(2), pp.312–320. Available at: <http://doi.wiley.com/10.1111/jam.13018> [Accessed 28 Apr. 2016].

Yamakawa, K., Kamiya, S., Nishida, S., Yoshimura, K., Yu, H., Lu, D. and Nakamura, S., 1988. Distribution of *Clostridium botulinum* in Japan and in Shinkiang District of China. *Microbiology and Immunology*, [online] 32(6), pp.579–587. Available at: <http://doi.wiley.com/10.1111/j.1348-0421.1988.tb01419.x> [Accessed 20 Oct. 2014].

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1), p.134.

Ye, Q., Roh, Y., Carroll, S.L., Blair, B., Zhou, J., Zhang, C.L. and Fields, M.W., 2004. Alkaline anaerobic respiration: isolation and characterization of a novel alkaliphilic and metal-reducing bacterium. *Applied and Environmental Microbiology*, [online] 70(9), pp.5595–602. Available at:

<http://aem.asm.org/content/70/9/5595.full> [Accessed 24 Feb. 2016].

Yutin, N. and Galperin, M.Y., 2013. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environmental Microbiology*, [online] 15(10), p.n/a-n/a. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4056668&tool=pmcentrez&rendertype=abstract> [Accessed 7 Apr. 2015].

Zeileis, A., Kleiber, C. and Jackman, S., 2008. Regression models for count data in R. *Journal of Statistical Software*, 27(8), pp.1–25.

Zhang, Q., He, X. and Yan, T., 2015. Differential decay of wastewater bacteria and change of microbial communities in beach sand and seawater microcosms. *Environmental science & technology*, 49(14), pp.8531–8540.

APPENDICES

Supplementary Table S3. Bacterial genomes sequences used in the study

Strain	Assembly level	Accession no.
Acetoanaerobium sticklandii str. DSM 519T	complete	NC_014614.1
Cellulosilyticum lentocellum DSM 5427^T	complete	NC_015275.1
Clostridioides difficile M120	complete	NC_017174.1
Clostridioides difficile 2007855	complete	NC_017178.1
Clostridioides difficile 630	complete	NZ_CP010905.1
Clostridioides difficile 630T	complete	NC_009089.1
Clostridioides difficile ATCC 43255	draft	NZ_CM000604.1
Clostridioides difficile ATCC 9689 = DSM 1296	complete	CP011968.1
Clostridioides difficile BI1	complete	NC_017179.1
Clostridioides difficile BI9	draft	NC_013974.1
Clostridioides difficile CD196	complete	NC_013315.1
Clostridioides difficile CD630DERM	complete	NZ_LN614756.1
Clostridioides difficile CF5	complete	NC_017173.1
Clostridioides difficile CIP 107932	contig	NZ_CM000659.1
Clostridioides difficile G46R2_Q2	contig	NZ_CDND01000001.1
Clostridioides difficile M68	draft	NC_017175.1
Clostridioides difficile QCD-23m63	contig	NZ_CM000660.1
Clostridioides difficile QCD-32g58	contig	NZ_CM000287.1
Clostridioides difficile QCD-37x79	contig	NZ_CM000658.1
Clostridioides difficile QCD-63q42	contig	NZ_CM000637.1
Clostridioides difficile QCD-66c26	contig	NZ_CM000441.1
Clostridioides difficile QCD-76w55	contig	NZ_CM000661.1
Clostridioides difficile QCD-97b34	contig	NZ_CM000657.1
Clostridioides difficile R20291	draft	NC_013316.1
Clostridioides mangenotii LM2T	contig	NZ_JIAA00000000.1
C. aceticum str. DSM 1496^T	complete	NZ_CP009687.1
C. acetobutylicum ATCC 824^T	complete	NC_003030.1
C. acetobutylicum DSM 1731	complete	NC_015687.1
C. acetobutylicum EA 2018	complete	NC_017295.1
C. algidicarnis B3^T	contig	NZ_JNLN01000001.1

Strain	Assembly level	Accession no.
C. autoethanogenum DSM 10061^T	complete	NC_022592.1
C. baratii str. Sullivan^T	complete	NZ_CP006905.1
C. beijerinckii ATCC 35702	complete	NZ_CP006777.1
C. beijerinckii NCIMB 14988	complete	NZ_CP010086.1
C. beijerinckii NCIMB 8052^T	complete	NC_009617.1
C. bormimense M2/40^T	complete	NZ_HG917868.1
C. botulinum 202F	complete	NZ_CP006903.1
C. botulinum A str. ATCC 19397	complete	NC_009697.1
C. botulinum A str. ATCC 3502^T	complete	NC_009495.1
C. botulinum A str. Hall^T	complete	NC_009698.1
C. botulinum A2 str. Kyoto	complete	NC_012563.1
C. botulinum A3 str. Loch Maree	complete	NC_010520.1
C. botulinum B str. Eklund 17B	complete	NC_010674.1
C. botulinum B str. Eklund 17B(NRP)	complete	NC_018648.1
C. botulinum B str. Osaka05	scaffold	NZ_DF384213.1
C. botulinum B1 str. Okra	complete	NC_010516.1
C. botulinum Ba4 str. 657	complete	NC_012658.1
C. botulinum CDC_1436	complete	NZ_CP006908.1
C. botulinum CDC_297	complete	CP006907.1
C. botulinum E1 str. 'BoNT E Beluga'	contig	NZ_ACSC01000002.1
C. botulinum E3 str. Alaska E43	complete	NC_010723.1
C. botulinum F str. 230613	complete	NC_017297.1
C. botulinum F str. Langeland	complete	NC_009699.1
C. botulinum H04402 065	complete	NC_017299.1
C. botulinum NCTC 8266	complete	NZ_CP010520.1
C. botulinum NCTC 8550	complete	NZ_CP010521.1
C. botulinum Prevot_594	complete	NZ_CP006902.1
C. botulinum str. 111	complete	NZ_AP014696.1
C. butyricum str. CDC_51208	complete	NZ_CP013239.1
C. carboxidivorans P7	complete	NZ_CP011803.1
C. cellulovorans 743B^T	complete	NC_014393.1
C. colicanis 209318	scaffold	NZ_KB850956.1
C. intestinale URNW	scaffold	NZ_KI273145.1
C. kluyveri DSM 555^T	complete	NC_009706.1
C. kluyveri NBRC 12016	complete	NC_011837.1
C. ljungdahlii DSM 13528	complete	NC_014328.1
C. novyi NT^T	complete	NC_008593.1
C. pasteurianum ATCC 6013	complete	NZ_CP009267.1

Strain	Assembly level	Accession no.
C. pasteurianum BC1^T	complete	NC_021182.1
C. pasteurianum DSM 525	complete	NZ_CP009268.1
C. pasteurianum NRRL B-598	complete	NZ_CP011966.1
C. perfringens ATCC 13124	complete	NC_008261.1
C. perfringens F262	draft	NZ_CM001477.1
C. perfringens str. 13^T	complete	NC_003366.1
C. perfringens str. FORC_003	complete	NZ_CP009557.1
C. saccharobutylicum DSM 13864^T	complete	NC_022571.1
C. saccharoperbutylacetonicum N1-4(HMT)^T	complete	NC_020291.1
C. scatologenes str. ATCC 25775^T	complete	NZ_CP009933.1
C. sporogenes str. DSM 795	complete	NZ_CP011663
C. sporogenes str. NCIMB 10696	complete	NZ_CP009225.1
C. tetani 12124569	complete	NC_022777.1
C. tetani E88^T	complete	NC_004557.1
C. botulinum BKT015925	complete	NC_015425.1
Gottschalkia acidurici 9a^T	complete	NC_018664.1
L. bolteae 90A9^T	scaffold	NZ_KB851182.1
L. hathewayi WAL-18680^T	scaffold	NZ_JH379027-44.1
L. indolis DSM 755^T	contig	NZ_AZUI01000001.1
L. phytofermentans ISDg^T	complete	NC_010001.1
L. phytofermentans KNHs212	contig	NZ_JNLM01000001.1
L. saccharolyticum WM1^T	complete	NC_014376.1
Paeniclostridium sordellii JGS6382	draft	LN681234.1
Paeniclostridium sordellii ATCC 9714	draft	LN679998.1
R. cellulolyticum H10^T	complete	NC_011898.1
R. cellulosi DG5^T	complete	NZ_LM995447.1
R. clariflavum DSM 19732^T	complete	NC_016627.1
R. josui JCM 17888^T	contig	NZ_JAGE01000001.1
R. stercorarium subsp. stercorarium DSM 8532	complete	NC_020887.2
R. stercorarium subsp. stercorarium DSM 8532^T	complete	NC_020134.1
R. thermocellum ATCC 27405^T	complete	NC_009012.1
R. thermocellum DSM 1313	complete	NC_017304.1

Supplementary Table 2. Selected attributes and diversity characteristics for the 50 housekeeping genes used in this study.

Protein name	Gene abbreviation	Category [1]	Sites	G+C (%)	*S*	S (%)	*H*	π	TD*	FLD*
Aspartyl-tRNA synthase	***aspS***	RNA met.	1752	34%	1342	77%	68	0.296	-0.333	-0.914
Cysteinyl-tRNA synthase	***cysS***	RNA met.	1382	33%	891	64%	67	0.284	-0.475	0.321
D-alanyl-alanine synthetase A	***ddl***	E & I met.	875	33%	602	69%	69	0.366	-0.173	0.353
Replicative DNA helicase	***dnaB***	DNA met.	1293	35%	936	72%	65	0.317	-0.113	0.428
DNA primase	***dnaG***	DNA met.	1718	31%	1030	60%	71	0.352	-0.334	-0.583
Hsp70 co-chaperone	***dnaJ***	Protein P, F & S	1070	37%	705	66%	67	0.316	-0.269	0.257
Chaperone Hsp70	***dnaK***	Protein P, F & S	1835	35%	1182	64%	67	0.239	-0.158	0.501
DNA polymerase III, β subunit	***dnaN***	DNA met.	1080	28%	794	74%	67	0.342	-0.353	0.169
GTP binding protein	***engA***	RNA met.	1307	34%	886	68%	66	0.273	-0.114	0.638
Enolase	***eno***	E & I met.	1274	36%	838	66%	69	0.249	-0.411	0.103
GTP binding protein	***era***	RNA met.	879	32%	625	71%	68	0.313	-0.190	0.062
Protein component of signal recognition particle	***ffh***	Protein P, F & S	1286	35%	789	61%	68	0.278	-0.174	0.332
Signal recognition particle receptor	***ftsY***	Protein P, F & S	911	34%	595	65%	67	0.290	-0.127	1.137
Cytoskeletal cell division protein	***ftsZ***	Cellular processes	946	37%	601	64%	65	0.262	-0.118	-0.097
Elongation factor G	***fusA***	RNA met.	2029	37%	1468	72%	66	0.268	-0.479	0.410
Glyceraldehyde-3-phosphate dehydrogenase	***gapA***	E & I met.	987	36%	698	71%	64	0.295	-0.195	0.788
Glutamyl-tRNA synthase	***gltX***	RNA met.	1623	35%	1002	62%	70	0.229	-0.288	0.069
Class I heat shock protein	***groEL***	Protein P, F & S	1438	33%	1011	70%	72	0.357	-0.078	1.373
DNA gyrase, A subunit	***gyrA***	DNA met.	2380	33%	1779	75%	71	0.290	-0.204	0.274
DNA gyrase, B subunit	***gyrB***	DNA met.	1862	34%	1227	66%	67	0.276	-0.132	0.664

Protein name	Gene abbreviation	Category [1]	Sites	G+C (%)	*S*	S (%)	*H*	π	TD*	FLD*
Initiation factor IF-2	***infB***	RNA met.	2040	35%	1255	62%	68	0.275	-0.276	0.572
ATP-dependent protease La	***lonA***	Protein P, F & S	2290	33%	1715	75%	74	0.326	-0.212	0.665
Cysteine desulfurase-NifS homolog	***iscS***	RNA met.	1129	34%	858	76%	67	0.356	-0.117	0.232
Dimethyladenosine transferase	***ksgA***	RNA met.	782	32%	591	76%	68	0.338	-0.197	0.483
GTP binding elongation factor	***lepA***	RNA met.	1785	35%	1147	64%	70	0.262	0.071	1.078
Leucyl-tRNA synthase	***leuS***	RNA met.	2430	35%	1541	63%	70	0.303	-0.262	0.809
Methionine aminopeptidase	***map1***	Protein P, F & S	746	37%	511	68%	66	0.371	0.184	1.402
Methionine adenosyltransferase	***metK***	E & I met.	1172	36%	782	67%	65	0.241	-0.523	-0.715
tRNA (5-methylaminomethyl-2-thiouridylate) methyl- transferase	***mnmA***	RNA met.	1038	33%	744	72%	67	0.301	-0.442	0.141
GTP binding protein involved in biosynthesis of 5-methilaminomethyl-2-thiouridine	***mnmE***	RNA met.	1369	31%	927	68%	67	0.327	-0.359	-0.690
S-adenosyl-methyltransferase	***mraW***	RNA met.	918	33%	670	73%	71	0.302	-0.311	1.151
Pyruvate-flavodoxin dehydrogenase	***nifJ***	E & I met.	3475	36%	2352	68%	71	0.282	-0.154	0.681
Transcription-translation coupling	***nusA***	RNA met.	1061	33%	739	70%	64	0.280	-0.441	0.133
GTP binding protein	***obgE***	RNA met.	1273	36%	927	73%	68	0.308	-0.197	-0.025
6-Phosphofructokinase	***pfkA***	E & I met.	941	37%	678	72%	65	0.294	-0.304	0.031
Glucose-6-phosphate isomerase	***pgi***	E & I met.	1324	34%	877	66%	68	0.262	-0.386	-0.299

Phosphoglycerate kinase	***pgk***	E & I met.	1187	35%	777	65%	69	0.246	-0.340	0.230
Phenylalanyl-tRNA synthase, a subunit	***pheS***	RNA met.	1020	35%	773	76%	66	0.290	-0.306	0.782
Phenylalanyl-tRNA synthase, b subunit	***pheT***	RNA met.	2365	33%	1646	70%	67	0.336	-0.376	0.755
Peptide chain release factor 1 (RF1)	***prfA***	RNA met.	1055	35%	770	73%	66	0.278	-0.246	0.299
50S ribosomal protein L2	***rplB***	RNA met.	815	38%	534	66%	62	0.241	-0.239	1.163
RNA polymerase, α subunit	***rpoA***	RNA met.	931	33%	631	68%	63	0.257	-0.133	0.267
RNA polymerase major σ factor	***rpoD***	RNA met.	1039	35%	661	64%	66	0.230	-0.352	1.535
Preprotein translocase subunit (ATPase)	***secA***	Protein P, F & S	2487	34%	1600	64%	68	0.278	-0.155	1.082
Membrane-embedded preprotein translocase subunit	***secY***	Protein P, F & S	1269	33%	860	68%	62	0.347	-0.117	-0.264
Seryl-tRNA synthase	***serS***	RNA met.	1261	33%	1005	80%	69	0.308	-0.558	-1.388
Threonyl-tRNA synthase	***thrS***	RNA met.	1889	34%	1254	66%	69	0.302	-0.192	-0.080
Thioredoxin reductase	***trxB***	E & I met.	833	34%	590	71%	68	0.339	-0.502	0.410
Elongation factor Ts	***tsf***	RNA met.	874	35%	438	50%	57	0.271	-0.224	0.213
GTP binding protein	***ychF***	RNA met.	1095	34%	748	68%	65	0.268	-0.144	0.552

Supplementary Table 3. All *in silico* primer-Blast results retrieved using AMRISA primer set

Accession	*Strain*	Dye	Peaks															
CP006905.1	*C. baratii str. Sullivan*	VIC	530	530	531	531	532	532	559	560	616	698						
CP014204.2	*C. baratii strain CDC51267*	VIC	532	532	533	533	533	533	534	559	616	699						
CP006777.1	*C. beijerinckii ATCC 35702*	VIC	499	499	499	499	499	499	499	579	617	625	625	625	687	795		
LN908213.1	*C. beijerinckii DSM 6423*	VIC	499	499	499	499	499	579	615	625	625	625	664	673	793	799	800	
CP000721.1	*C. beijerinckii NCIMB 8052*	VIC	499	499	499	499	499	499	499	579	617	625	625	625	687	795		
CP011966.2	*C. beijerinckii NRRL B-598*	VIC	499	499	578	615	624	624	624	624	624	624	624	624	690	749	795	800
CP016090.1	*C. beijerinckii strain BAS/B3/I/124*	VIC	498	498	499	499	579	617	623	624	624	687	786	794	795	799		
CP010086.2	*C. beijerinckii strain NCIMB 14988*	VIC	498	499	499	578	616	623	623	624	624	624	624	624	690	785	786	861
CP006903.1	*C. botulinum 202F*	VIC	504	504	504	505	505	576	623	624	667	863						
FR745875.1	*C. botulinum B str. Eklund 17B(NRP)*	VIC	476	505	505	505	505	505	665	862								
CP001056.1	*C. botulinum B str. Eklund 17B*	VIC	476	505	505	505	505	505	505	505	505	665	862					
CP001078.1	*C. botulinum E3 str. Alaska E43*	VIC	506	506	506	506	506	506	506	506	506	656	665					
CP010520.1	*C. botulinum strain NCTC 8266*	VIC	506	506	506	506	506	506	506	506	506	656	665					
CP010521.1	*C. botulinum strain NCTC 8550*	VIC	506	506	506	506	506	506	506	506	506	656	665					
CP013239.1	*C. butyricum strain CDC_51208*	VIC	518	519	520	520	521	521	521	607	633	642	685	840				
CP013352.1	*C. butyricum strain JKY6D1*	VIC	518	518	518	569	604	631	642	642	685							
CP013252.1	*C. butyricum strain KNU-L09*	VIC	518	519	519	519	569	605	631	642	643	685	1083					
CP016332.1	*C. butyricum strain TK520*	VIC	685	685	685	685	685	686	686	686	686							
CP014704.1	*C. butyricum strain TOA*	VIC	517	518	519	519	519	642	718	1081								
CP011803.1	*C. carboxidivorans P7*	VIC	499	499	500	500	500	500	500	500	500	601	672					
LT799839.1	*C. chauvoei JF4335*	VIC	507	507	507	507	507	507	507	584	662							

CP018630.1	*C. chauvoei strain 12S0467*	VIC	507	507	507	507	507	507	507	584	662			
CP018624.1	*C. chauvoei strain DSM 7528*	VIC	507	507	507	507	507	507	507	584	662			
CP020953.1	*C. drakei strain SL1*	VIC	499	499	499	499	499	499	573	663	677	846		
CP000246.1	*C. perfringens ATCC 13124*	VIC	464	464	464	464	464	464	697	697				
AP017630.1	*C. perfringens DNA*	VIC	464	464	464	464	464	464	464	464	697	697		
CP000312.1	*C. perfringens SM101*	VIC	464	464	464	464	464	464	464	464	697	698		
BA000016.3	*C. perfringens str. 13 DNA*	VIC	464	464	464	464	464	696	696	698	698			
CP019468.1	*C. perfringens strain CP15*	VIC	463	463	463	463	463	464	464	464	696	697		
CP019576.1	*C. perfringens strain Del1*	VIC	463	464	464	464	464	464	464	464	695	697		
CP025501.1	*C. perfringens strain EHE-NE18*	VIC	464	464	464	464	464	464	464	464	697	697		
CP009557.1	*C. perfringens strain FORC_003*	VIC	464	464	464	464	464	464	464	464	696	697		
CP013101.1	*C. perfringens strain FORC_025*	VIC	464	464	464	464	464	464	464	464	696	698		
CP010993.1	*C. perfringens strain JP55*	VIC	464	464	464	464	464	464	464	464	697	697		
CP010994.1	*C. perfringens strain JP838*	VIC	464	464	464	464	464	464	464	464	697	698		
CP028149.1	*C. perfringens strain JXJA17*	VIC	464	464	464	464	464	464	464	464	696	697		
CP023410.1	*C. perfringens strain LLY_N11*	VIC	464	464	464	464	464	464	464	464	695	697		
LS483393.1	*C. perfringens strain NCTC 13170*	VIC	464	464	464	464	464	464	464	464	697	697		
LS483461.1	*C. perfringens strain NCTC 2837*	VIC	464	464	464	464	464	464	464	464	696	696		
CP006721.1	*C. saccharobutylicum DSM 13864*	VIC	520	520	520	520	520	555	555	605	621	633	683	724
CP016089.1	*C. saccharobutylicum strain BAS/B3/SW*	VIC	520	520	520	520	520	555	555	605	621	633	683	724
CP016092.1	*C. saccharobutylicum strain NCP 195*	VIC	520	520	520	520	520	555	555	605	621	633	683	724
CP016086.1	*C. saccharobutylicum strain NCP 200*	VIC	520	520	520	520	520	555	555	605	621	633	683	724
CP016091.1	*C. saccharobutylicum strain NCP 258*	VIC	520	520	520	520	520	555	555	605	621	633	683	708
CP009933.1	*C. scatologenes strain ATCC 25775*	VIC	499	499	499	499	499	499	499	499	773			

CP014331.1	*C. sp. MF28*	VIC	499	499	499	578	615	623	624	624	624	624	624	624	625	690	787	793
CP017253.2	*C. taeniosporum strain 1/k*	VIC	502	503	621	664	745	784	787									
CP016280.1	*C. tyrobutyricum strain W428*	VIC	517	517	517	517	517											
CP000726.1	*C. botulinum A str. ATCC 19397*	VIC	344	344	345	345	345	345	360	498								
AM412317.1	*C. botulinum A str. ATCC 3502*	VIC	344	344	345	345	345	345	360	498	498							
CP000727.1	*C. botulinum A str. Hall*	VIC	344	345	345	345	345	345	360	498								
CP001581.1	*C. botulinum A2 str. Kyoto*	VIC	281	345	345	345	345	350	350	360	498							
CP000962.1	*C. botulinum A3 str. Loch Maree*	VIC	345	345	345	345	345	345	345	360	498							
CP000939.1	*C. botulinum B1 str. Okra*	VIC	283	283	283	347	347	497										
CP001083.1	*C. botulinum Ba4 str. 657*	VIC	498															
CP006908.1	*C. botulinum CDC_1436*	VIC	498															
AP014696.1	*C. botulinum DNA*	VIC	344	344	344	345	345	345	345	345	360	498						
CP002011.1	*C. botulinum F str. 230613*	VIC	345	345	345	345	345	345	345	362	558							
CP000728.1	*C. botulinum F str. Langeland*	VIC	281	345	345	345	362	558										
FR773526.1	*C. botulinum H04402 065*	VIC	345	345	345	345	345	345	346	360								
CP006902.1	*C. botulinum Prevot_594*	VIC	350	350	350	350	559											
CP013681.1	*C. botulinum strain 1169*	VIC	345	345	348	281	345	350	345	360	498							
CP013845.1	*C. botulinum strain A634*	VIC	498			345												
CP013847.1	*C. botulinum strain Af650*	VIC	345	345	350	281	345	350	345	360	498							
CP013841.1	*C. botulinum strain AM1051*	VIC	345	345	348	283	346	360	348	498								
CP013701.1	*C. botulinum strain AM1195*	VIC	347	347	351	347	350	351	351	420	559							
CP013683.1	*C. botulinum strain AM282*	VIC	345	345	348	283	345	348	345	360	498							
CP013850.1	*C. botulinum strain B305*	VIC	344	344	345	344	344	345	345	360	498							
CP014219.1	*C. botulinum strain B515*	VIC	344	344	345	344	344	345	345	360	498							

CP014174.1	*C. botulinum strain B609*	VIC	344	344	345	344	344	345	345	360	498			
CP013843.1	*C. botulinum strain B742*	VIC	344	344	345	344	345	345	345	360	498			
CP014151.1	*C. botulinum strain BrDura*	VIC	283	283	345	283	283	348	345	498				
CP013243.1	*C. botulinum strain CDC_1632*	VIC	347	348	348	348	405	405	420	559				
CP013247.1	*C. botulinum strain CDC_53174*	VIC	283	283	283	345	345	345	347	360	559			
CP013296.1	*C. botulinum strain CDC_54064*	VIC	345	345	350	281	345	360	348	498	498			
CP013242.1	*C. botulinum strain CDC_67071*	VIC	347	347	347	347	347	348	411	424	561			
CP014148.1	*C. botulinum strain CDC_67190*	VIC	345	345	345	345	345	350	345	360	497			
CP013246.1	*C. botulinum strain CDC_69094*	VIC	345	345	346	343	345	350	345	360	498			
CP013857.1	*C. botulinum strain CDC_69096*	VIC	345	498										
CP031098.1	*C. botulinum strain CFSAN034202*	VIC	345	345	345	345	345	345	345	497	497			
CP028859.1	*C. botulinum strain CFSAN064329*	VIC	344	344	344	345	345	345	345	360	498			
CP028842.1	*C. botulinum strain DFPST0029*	VIC	344	345	344	345	344	345	345	360	498			
CP013686.1	*C. botulinum strain F1425*	VIC	344	345	345	344	345	345	345	360	498			
CP013705.1	*C. botulinum strain F160*	VIC	344	344	344	344	345	345	345	359	498			
CP013849.1	*C. botulinum strain Mauritius*	VIC	283	283	345	283	345	347	345	360	559			
CP027780.1	*C. botulinum strain Mfbjulcb3*	VIC	347	347	351	347	350	351	351	420	559			
CP011803.1	*C. carboxidivorans P7*	VIC	349	349	350	350	350	350	350	350	350	451	522	
CP020953.1	*C. drakei strain SL1*	VIC	349	349	423	696								
CP009933.1	*C. scatologenes strain ATCC 25775*	VIC	349	349	349	349	349	349	349	349	623			
EU333998.1	*C. sporogenes clone CMCC64941*	VIC	350											
CP011663.1	*C. sporogenes strain DSM 795*	VIC	349	350	350	350	350	350	350	422	559			
CP009225.1	*C. sporogenes strain NCIMB 10696*	VIC	349	350	350	350	350	350	350	422	559			
CP002410.1	*C. botulinum BKT015925*	NED	355	355	355	355	355	355	355	355	446	499		

CP000382.1	*C. novyi NT*	NED	292	292	292	292	292	292	292	292	484	529		
LT906477.1	*C. cochlearium strain NCTC13027*	PET	507	507	507	507	507	589	773					
HG530135.1	*C. tetani 12124569 main*	PET	504	504	504	588	588							
AE015927.1	*C. tetani E88*	PET	504	504	504	588	588	787						
CP010905.2	*Clostridioides difficile 630*	6FAM	577	617	617	639	639	666	737	741	798	859	859	860
CP016318.1	*Cl. difficile strain 630 delta erm*	6FAM	577	617	617	639	639	666	737	741	798	859	859	860
AM180355.1	*Cl. difficile 630*	6FAM	577	615	617	639	666	737	741	798	859	859	860	
CP011847.1	*Cl. difficile strain DSM 27639*	6FAM	577	617	617	617	639	666	737	741	798	859	859	860
CP014150.1	*Cl. difficile CD630DERM*	6FAM	577	617	617	639	639	666	737	741	798	859	859	860
FN538970.1	*Cl. difficile CD196*	6FAM	542	575	637	637	639	656	669	678	735	859		
FN545816.1	*Cl. difficile R20291*	6FAM	542	575	575	637	637	637	639	678	735	859		
CP003939.1	*Cl. difficile BJ08*	6FAM	542	574	574	735	735	736	859					
CP026594.1	*Cl. difficile strain 08-00495*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859
CP010888.1	*Cl. difficile strain 08ACD0030*	6FAM	542	544	575	617	617	637	639	796	797	859	859	
CP026599.1	*Cl. difficile strain 09-00072*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859
CP026596.1	*Cl. difficile strain 10-00071*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859
CP026597.1	*Cl. difficile strain 10-00078*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859
CP026598.1	*Cl. difficile strain 10-00253*	6FAM	542	575	575	637	639	678	735	735	735	757	757	859
CP026593.1	*Cl. difficile strain 12-00008*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859
CP026595.1	*Cl. difficile strain 12-00011*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859
CP027014.1	*Cl. difficile strain AK*	6FAM	542	542	542	575	637	637	639	795	798	798	798	857
CP019870.1	*Cl. difficile strain BR81*	6FAM	575	595	595	595	637	637	637	756	797	859	861	861
CP026592.1	*Cl. difficile strain CD-10-00484*	6FAM	542	575	575	637	637	639	678	735	757	757	859	859
CP029154.1	*Cl. difficile strain CD161*	6FAM	541	542	575	619	678	679	679	734	734	736	736	861

CP026591.1	*Cl. difficile strain CD-17-01474*	6FAM	542	575	575	637	637	637	639	678	735	735	757	859					
CP029152.1	*Cl. difficile strain CDT4*	6FAM	541	542	575	619	678	679	679	734	734	736	736	861					
CP022524.1	*Cl. difficile strain DH/NAP11/ST-42*	6FAM	575	595	595	595	637	637	637	756	797	859	861	861					
CP011846.1	*Cl. difficile strain DSM 27638*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859					
CP011968.1	*Cl. difficile ATCC 9689 = DSM 1296*	6FAM	542	542	542	575	637	637	639	795	798	798	857						
CP011848.1	*Cl. difficile strain DSM 27640*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859					
CP020424.2	*Cl. difficile strain FDAARGOS_267*	6FAM	542	542	542	575	637	637	639	795	798	798	857						
CP019469.1	*Cl. difficile strain LEM1*	6FAM	542	544	575	617	617	637	796	796	859	859							
CP025044.1	*Cl. difficile strain R0104a*	6FAM	542	575	575	637	637	639	678	735	735	757	757	859					
CP026613.2	*Cl. difficile strain R1*	6FAM	619	637	639	679	681	681	723	757	757	757	799	799					
CP026614.2	*Cl. difficile strain R2*	6FAM	619	637	639	679	681	681	723	757	757	757	799	799					
CP026615.2	*Cl. difficile strain R3*	6FAM	619	637	639	679	681	681	723	757	757	757	799	799					
CP025047.1	*Cl. difficile strain W0003a*	6FAM	595	595	637	637	637	637	639	756	756	797	859	861					
CP025046.1	*Cl. difficile strain W0022a*	6FAM	542	544	575	617	617	637	796	796	859	859	859						
CP025045.1	*Cl. difficile strain W0023a*	6FAM	575	595	595	595	637	637	637	756	797	859	861	861					
CP013196.1	*Cl. difficile strain Z31*	6FAM	693	694	694	694	694	694	695	695	695	697							
FN668941.1	*Cl. difficile BI1*	6FAM	542	575	575	637	637	639	678	735	757	757	757	859					
FN665654.1	*Cl. difficile*	6FAM	542	575	575	617	637	639	678	757	859								
FN665652.1	*Cl. difficile*	6FAM	528	542	575	619	678	679	679	734	736								
FN665653.1	*Cl. difficile*	6FAM	617	618	619	639	681	681	723	757	757	798	799						
FN668375.1	*Cl. difficile M68*	6FAM	542	542	575	575	575	619	619	678	678	679	679	679	679	679	734	736	
LN614756.1	*Paeniclostridium sordellii*	6FAM	496	553	555	555	555	555	555	555	555	555	603	603	603	603	648	656	
LN555523.1	*Romboutsia ilealis strain CRIB*	6FAM	499	499	579	579	579	579	634	634	641	641	646	696	696	696			
LN650648.1	*Romboutsia sp. Frifi strain FRIFI*	6FAM	495	558	558	558	558	586	586	589	613	621	621	621	623	691	696	696	

Supplementary Table 4A. Estimated coefficients of measured predictors on the relative abundance of selected 6FAM peaks. Negative binomial part of models. **Model:** Full model used (NB: negative binomial, hurdle: NBH). **Zero counts:** number of sites where peak absent. **Baseline RA:** model-predicted baseline relative abundance for arable, dry hotspot in August 2018. Significant predictors denoted with Asterix (p<0.05 = *, p<0.01 =**, p<0.001 =***). Colouring used to indicate scale of predictor of RA, red: increases, blue: decreases.

Predictor/Peak	***493***	***546***	***576***	***595***	***531***	***625***	***520***	***524***	***579***	***603***	***700***
Model	NB	NBH	NBH	NBH	NBH	NBH	NBH	NBH	NBH	NBH	NBH
Zero counts:	**7**	**60**	**77**	**165**	**57**	**112**	**214**	**350**	**384**	**389**	**416**
Baseline RA (%)	12.4 ***	1.6 ***	1.5 ***	1.3 ***	1.1 ***	1.3 ***	2.4 ***	0.75 ***	1.09	0.85 ***	0.8 ***
Relative soil water						1.01 ***					1.04
HarvestMay.18	0.97	0.91 *	0.95	0.82 .	1.00	1.01	1.36 *	0.84		2.01	1.23
HarvestFeb.18	1.17 *	1.08	1.15 **	1.21	1.60 **	1.35 .	0.90	1.42		0.77 *	0.00
HarvestNov.17	1.04	1.01	0.92 .	0.84	1.35 **	1.07	1.01	1.28 *		0.79 .	0.75
HarvestSept.17	1.03	0.99	1.02	1.00	1.22 .	0.79	1.46 **	1.11		0.85	0.55 **
Beef farm	1.16 .	1.45 ***	1.51 ***	1.10	1.87 ***	1.27 *	0.41 ***	1.40 .		1.01	1.68 *
Dairy Farm	1.17 *	1.53 ***	1.16 **	1.26 *	1.03	0.86	0.23 ***	1.02		0.51 ***	0.93
Mixed Grazing	0.97	1.43 ***	1.16 *	0.70 *	2.29 ***	2.54 ***	0.20 ***	0.89		0.47 **	0.67
Site- wetspot	0.88 **	0.82 ***	0.90 ***	0.94 .	1.01	0.85 **	1.92 ***	1.29 **		1.61 ***	1.46 .
Grazing Intensity	0.97	0.91 ***	0.93 ***	0.99	0.94 .		1.12 *	0.94		0.94	
Wetspot x Grazing		1.06 *			1.02						1.14
Feb.18 x Beef				0.94	0.57 **	0.82					1.00
May.18 x Beef				1.12	0.96	1.53 **					0.95
Nov.17 x Beef				1.00 *	0.75 .	0.76					0.98
Sept.17 x Beef				0.70	0.65 **	1.18					
Feb.18 x Dairy				0.92	0.83	1.10					
May.18 x Dairy				1.01	0.88	0.87					
Nov.17 x Dairy				0.95	0.68 **	0.53 ***					
Sept.17 x Dairy				0.92	1.01	1.00					
Feb.18 x Mixed Grazing				0.89	0.80	0.89					
May.18 x Mixed Grazing				1.06	1.08	1.50 *					
Nov.17 x Mixed Grazing				0.81	0.92	0.93					
Sept.17 x Mixed Grazing				0.87	1.12	1.04					

Supplementary Table 4B. Estimated coefficients of predictors for binary part of hurdle models on selected 6FAM peaks. Binary (zero-count) part of models. Coefficients indicate the change in probability of the peak being detected at a site, having been exponentiated. **Model:** Main overall model used (Hurdle: NBH) **Baseline RA:** model-predicted baseline relative abundance for arable, dry hotspot in August 2018. Significant predictors denoted with Asterix ($p<0.05$ = *, $p<0.01$ =**, $p<0.001$ =***). Colouring used to indicate scale of predictor of RA, red: increases, blue: decreases.

Predictor/Peak	***546***		***576***		***595***		***531***		***625***		***520***		***524***		***579***		***603***		***700***	
Model	NBH		NBH		NBH		NBH		NBH		NBH		NBH		NBH		NBH		NBH	
Zero counts:																				
Baseline RA (%)	222	***	11.2	***	7.23	***	85	***	17.58	***	0.27	***	1.54		0.1	***	0.37		0.1	***
Relative soil water									1.01										1.01	
HarvestMay.18	1.38		2.03		1.84		1.08		1.20E+07		0.37	*	1.43		0.13	***	0.03	***	2.26	
HarvestFeb.18	0.19	**	0.39	*	0.22	*	0.02	***	0.03	***	1.33		0.06	***	0.98		1.67		0.00	.
HarvestNov.17	0.34	*	0.49	.	1.01		0.26		0.17	*	0.62		0.72		0.65		0.93		1.01	
HarvestSept.17	0.45		0.72		0.66		0.09	*	0.04	***	0.82		0.97		0.66		0.94		1.19	
Beef farm	8.05	***	18.16	***	15.44	**	5.18		1.95		0.00	***	0.02	***	0.28	*	0.03	***	0.18	***
Dairy Farm	10.77	***	13.02	***	15.85	**	6.12E+07		0.46		0.02	***	0.06	***	0.27	**	0.07	***	0.10	***
Mixed Grazing	19.94	***	5.58	***	0.97		7.37		1.08		0.01	***	0.02	***	0.28	**	0.01	***	0.03	***
Site- wetspot	0.02	***	0.23	***	0.18	***	0.17	*	0.46	*	3.18	***	1.40		5.54	***	4.57	***	7.99	***
Grazing Intensity	0.14	***	0.55	**	0.52	**	0.47				1.45	*	1.88	***	1.41	*	1.71	**		
Wetspot x Grazing	4.63	**													Feb.18 x r. soil water				1.18	
Feb.18 x Beef					1.53				10.73						May.18 x r. Soil water				1.01	
May.18 x Beef					0.48				0.00						Nov.17 x Mixed Grazing				0.98	
Nov.17 x Beef					0.59				1.32						Sept.17 x Mixed Grazing				0.89	*
Sept.17 x Beef					0.65				5.59											
Feb.18 x Dairy					0.73				1.17	*										
May.18 x Dairy					0.31				0.00											
Nov.17 x Dairy					1.62				30.33	*										
Sept.17 x Dairy					3.53				1.02											
Feb.18 x Mixed Grazing					0.50				28.73	.										
May.18 x Mixed Grazing					43.69	**			1.01	*										
Nov.17 x Mixed Grazing					0.04	*			12.86											
Sept.17 x Mixed Grazing					2.14				12.28											

Supplementary Table 5A. Estimated coefficients of measured predictors on the relative abundance of selected VIC peaks.
Model: Full model used (hurdle: NBH). Zero counts: number of sites where peak absent. Baseline RA: model-predicted baseline relative abundance for arable, dry hotspot in August 2018. Significant predictors denoted with Asterix (p<0.05 = *, p<0.01 =**, p<0.001 =***). Colouring used to indicate scale of predictor of RA, red: increases, blue: decreases.

Predictor/Peak	***340***	***346***	***416***	***491***	***496***	***459***	***504***	***529***
Model	NBH	NBH	NBH	NBH	NBH	NBH	NBH	NBH
Zero counts:	**324**	**163**	**399**	**92**	**202**	**419**	**422**	**318**
Baseline RA (%)	0.55 ***	1.61 ***	0.83 ***	2.56 ***	1.21 ***	0.67 ***	0.53 ***	0.61 ***
Relative soil water				0.99 **	1.02 ***			0.99 .
HarvestMay.18		0.89	0.32 ***	0.68 **	1.01		1.73 *	0.79
HarvestFeb.18		1.26	0.73	1.26	1.31 .		4.24 ***	1.66 **
HarvestNov.17		0.94	0.84	0.47 ***	0.63 **		1.17	0.79 .
HarvestSept.17		2.26 ***	0.66	0.57 ***	0.73 *		1.41	0.58 ***
Beef farm	1.39 *	0.91	2.27 **	0.50 ***	1.49 .			0.81
Dairy Farm	1.67 ***	1.06	1.00	3.42 ***	1.10			0.64 *
Mixed Grazing	1.04	0.96	1.42	0.91	2.08 ***			0.83
Site- wetspot				0.74 ***	0.65 ***			1.55 **
Grazing Intensity				0.82 ***	0.76 ***			1.20 ***
Wetspot x Grazing								3.65 ***
Feb.18 x Beef				1.18	0.67			
May.18 x Beef				1.21	0.98			
Nov.17 x Beef				1.98 **	0.95			
Sept.17 x Beef				2.99 ***	0.74			
Feb.18 x Dairy				0.57 *	0.43 **			
May.18 x Dairy				1.24	0.71			
Nov.17 x Dairy				1.58 *	0.71			
Sept.17 x Dairy				0.98	0.78			
Feb.18 x Mixed Grazing				1.13	0.98			
May.18 x Mixed Grazing				1.41	0.41 ***			
Nov.17 x Mixed Grazing				1.45	0.36 ***			
Sept.17 x Mixed Grazing				0.81	0.52 *			
			Site.TypeH:Grazing		1.20 **			
			Grazing:Water.Conc		0.99 **			

Supplementary Table 5B. Estimated coefficients of predictors for binary part of hurdle models on selected VIC peaks

Predictor/Peak	***340***	***346***	***416***	***491***	***496***	***459***	***504***	***529***
Model								
Zero counts:								
Baseline RA (%)	0.002 ***	0.003 ***	0.001 ***	0.36 ***	0.10 ***	0.0003 ***	0.0008 ***	0.0043 .
Relative soil water			1.05 ***					1.05 ***
HarvestMay.18	1.29	2.30 **	0.93	1.85	0.92	1.01		3.00 **
HarvestFeb.18	0.66	2.02 *	0.35 *	0.27 **	0.22 ***	0.28		0.11 ***
HarvestNov.17	3.30 **	8.42 ***	1.24	0.24 **	0.35 **	3.07 *		0.17 ***
HarvestSept.17	3.32 **	48.05 ***	0.69	0.34 *	0.22 ***	4.48 **		0.30 **
Beef farm	9.25 ***	4.92 ***	4.27 **	0.29 ***	0.75	4.98 ***		0.31 **
Dairy Farm	3.98 ***	1.30	2.54 *	20.51 **	0.39 **	0.89		0.46 *
Mixed Grazing	0.75	1.27	1.82	0.41 *	0.22 ***	3.14 *		2.51 *
Site- wetspot	0.56 *			0.29 ***	0.57 **		1.33	3.09 **
Grazing Intensity							1.75 **	1.60 *
Wetspot x Grazing								0.65 .
Feb.18 x Beef								
May.18 x Beef								
Nov.17 x Beef								
Sept.17 x Beef								
Feb.18 x Dairy								
May.18 x Dairy								
Nov.17 x Dairy								
Sept.17 x Dairy								
Feb.18 x Mixed Grazing								
May.18 x Mixed Grazing								
Nov.17 x Mixed Grazing								
Sept.17 x Mixed Grazing								

Binary (zero-count) part of models. Coefficients indicate the change in probability of the peak being detected at a site, having been exponentiated. **Model:** Main overall model used (hurdle: NBH) **Baseline RA:** model-predicted baseline relative abundance for arable, dry hotspot in August 2018. Significant predictors denoted with Asterix ($p<0.05$ = *, $p<0.01$ =**, $p<0.001$ =***). Colouring used to indicate scale of predictor of RA, red: increases, blue: decreases.

Supplementary Table 6. Estimated coefficients of predictors for absence and binary parts of hurle models for selected NED peaks

Predictor/Peak	***344***		***403***		***424***	
Model	NBH		NBH		NBH	
Zero counts:	**165**		**115**		**66**	
Baseline RA (%)	3.14	***	2.24	***	5.52	***
Relative soil water						
HarvestMay.18	0.90		1.80	***	1.20	*
HarvestFeb.18	2.16	***	1.72	***	1.73	***
HarvestNov.17	1.24	*	1.35	***	1.12	
HarvestSept.17	0.96		1.10		0.88	.
Beef farm	0.85		1.06		1.12	
Dairy Farm	1.06		1.34	***	0.80	**
Mixed Grazing	1.73	***	1.97	***	0.90	
Site- wetspot	0.77	*			1.12	*
Grazing Intensity	0.94					
Wetspot x Grazing	1.24	**				
Binary part						
Baseline RA (%)	0.46	***	0.77	***	0.71	***
Relative soil water	1.08	***	1.03	**	1.03	**
HarvestMay.18	0.02	***	0.05	**	0.36	
HarvestFeb.18	0.01	***	0.00	***	0.04	***
HarvestNov.17	0.09	***	0.04	**	0.20	*
HarvestSept.17	0.30	*	0.17		0.79	
Beef farm			0.84		0.40	*
Dairy Farm			2.98	**	0.58	
Mixed Grazing			1.46		0.29	**
Site- wetspot						
Grazing Intensity	0.75	*				
Wetspot x Grazing						
Soil water x Grazing	0.97	***				

	B.484	B.493	B.517	B.520	B.524	B.531	B.540	B.546	B.554	B.572	B.576	B.579	B.585	B.595	B.603	B.608	B.625	B.631	B.700
B.484	1																		
B.493	0.059	1																	
B.517	-0.083	-0.441	1																
B.520	-0.133	-0.324	0.398	1															
B.524	-0.108	-0.464	0.435	0.619	1														
B.531	-0.162	-0.163	-0.060	-0.420	-0.183	1													
B.540	-0.107	-0.136	-0.104	-0.288	-0.173	0.377	1												
B.546	0.250	0.186	-0.259	-0.450	-0.334	0.292	-0.003	1											
B.554	0.100	-0.323	0.096	0.036	0.080	0.169	0.111	0.120	1										
B.572	0.205	-0.379	0.285	0.406	0.489	-0.188	-0.064	-0.178	0.167	1									
B.576	0.105	0.432	-0.388	-0.495	-0.357	0.179	0.229	0.355	-0.003	-0.264	1								
B.579	-0.101	-0.453	0.598	0.434	0.508	-0.073	-0.104	-0.258	0.086	0.284	-0.429	1							
B.585	0.388	-0.132	-0.045	-0.240	-0.200	0.212	0.344	0.183	0.250	0.099	0.066	-0.108	1						
B.595	0.307	0.427	-0.310	-0.193	-0.185	-0.190	-0.048	0.293	0.001	0.029	0.500	-0.296	0.092	1					
B.603	-0.103	-0.492	0.456	0.581	0.635	-0.223	-0.185	-0.348	0.136	0.451	-0.350	0.502	-0.206	-0.169	1				
B.608	0.681	-0.030	-0.032	-0.111	-0.071	-0.144	-0.088	0.222	0.183	0.255	0.045	-0.077	0.401	0.271	-0.039	1			
B.625	-0.122	-0.277	0.043	-0.338	-0.181	0.628	0.433	0.201	0.210	-0.073	0.109	-0.006	0.290	-0.275	-0.156	-0.040	1		
B.631	0.246	-0.074	-0.038	-0.293	-0.173	0.158	0.184	0.267	0.294	0.131	0.156	-0.095	0.413	0.179	-0.136	0.305	0.317	1	
B.700	-0.111	-0.477	0.445	0.508	0.541	-0.182	-0.156	-0.333	0.130	0.378	-0.327	0.495	-0.181	-0.220	0.768	-0.087	-0.101	-0.120	1

Supplementary Table 7. Spearman's rank correlation coefficients for all 6FAM-labelled peaks. Coefficients underlined have two-tailed significance values $p < 0.05$

	Y.327	Y.344	Y.365	Y.385	Y.397	Y.403	Y.406	Y.418	Y.424	Y.428	Y.431	Y.441	Y.443	Y.446	Y.493	Y.571
Y.327	1															
Y.344	0.202	1														
Y.365	0.194	0.654	1													
Y.385	0.015	0.096	0.098	1												
Y.397	-0.314	-0.024	-0.011	-0.051	1											
Y.403	0.187	0.256	0.200	0.023	-0.090	1										
Y.406	-0.160	0.043	0.093	-0.012	0.203	-0.086	1									
Y.418	-0.251	0.096	0.169	-0.026	0.629	-0.051	0.317	1								
Y.424	-0.037	-0.014	-0.068	-0.127	0.157	0.157	-0.011	0.114	1							
Y.428	-0.096	0.159	0.208	0.008	0.228	-0.015	0.360	0.325	-0.010	1						
Y.431	0.004	0.119	0.164	0.086	0.163	-0.005	0.222	0.206	-0.105	0.310	1					
Y.441	-0.150	0.074	0.171	0.003	0.438	-0.036	0.243	0.512	-0.027	0.264	0.252	1				
Y.443	0.295	0.212	0.169	0.003	-0.274	0.216	-0.039	-0.080	0.094	0.041	0.008	-0.117	1			
Y.446	-0.203	0.217	0.242	0.191	0.401	0.115	0.212	0.439	-0.014	0.271	0.248	0.352	-0.196	1		
Y.493	-0.012	-0.090	-0.045	0.200	-0.173	-0.011	-0.041	-0.111	-0.138	-0.034	-0.008	-0.086	0.030	-0.002	1	
Y.571	0.029	0.134	0.193	0.029	0.212	0.071	0.116	0.326	0.028	0.179	0.164	0.315	0.144	0.233	-0.030	1

Supplementary Table 8. Spearman's rank correlation coefficients for all NED-labelled peaks. Coefficients underlined have two-tailed significance values $p < 0.05$

The end.