IDENTIFICATION AND DIVERSITY OF MIXED STRAIN INFECTIONS OF MYCOBACTERIUM AVIUM SUBSPECIES PARATUBERCULOSIS IN DAIRY CATTLE FROM CANADIAN HERDS

by

© Alexander Stephen Byrne

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Memorial University of Newfoundland

November 2023

St. John’s, Newfoundland and Labrador

Abstract

Mycobacterium avium subsp. *paratuberculosis* (MAP) is the pathogen responsible for Johne's Disease (JD) in ruminants, a cause of substantial economic loss in the global dairy industry. JD is a major problem in dairy animals, and concerns have been raised regarding the association of MAP with Crohn's Disease in humans. Mixed strain infections (MSI) refer to the concurrent infection of a susceptible host with multiple strains of a single pathogenic species. Microevolution refers to within-host changes in an infecting strain which lead to genetically distinguishable progeny. Both processes influence host-pathogen dynamics and disease progression, but not much is known about their prevalence and impact on JD.

In this thesis we first conducted a systematic review on how MSIs are defined in the literature, how widespread the phenomenon is among host species and how MSIs are detected. Next, we characterized MAP isolates from 67 cows using whole genome sequencing (WGS) based methods. SNP-based analysis of WGS data, when combined with animal movement data, was able to document disease transmission between herds. Finally, up to 10 MAP isolates from 14 high-shedding animals were examined using WGS to identify the prevalence of MSIs and microevolution. Two animals showed evidence of microevolution, while the remaining twelve showed evidence of both microevolution and MSIs. Our results showed a need for further investigation of MSIs within animal-based infections, that examination of short sequence repeats allows for greater discrimination between strains, and that differences in these repeats significantly alter protein structure.

General Summary

Mycobacterium avium subsp. *paratuberculosis* (MAP) is the pathogenic bacteria which causes Johne's Disease (JD) in cattle and results in substantial economic loss in the global dairy industry. JD is a major problem in dairy animals, with concerns also raised regarding the association of MAP with a selection of human diseases. Mixed strain infections (MSI) refer to the simultaneous infection of an organism with multiple strains of a single pathogenic species. Microevolution refers to the slight changes of a strain that occur within an infected host, which results in genetically diverse strains over time. Both processes influence interactions between the host and pathogen, but not much is known about their prevalence and impact on JD.

To determine how MSIs in mycobacteria are defined, along with how common they are, what species they are found in, and how they can be detected, a review of the literature was conducted. 121 articles reported that mycobacterial MSIs could be found in both humans and animals, though most of the work was focused on bacteria more closely related to *Mycobacterium tuberculosis* (78.5%) than more distantly related mycobacteria (21.5%).

To gain an understanding of the variation present in different herds of MAP, whole genome sequencing was performed on MAP isolates grown from a selection of cows in 20 herds from Quebec and Ontario. Analysis using two methods, MIRU-VNTR and ML-SSR, showed that the latter was able to identify distinct strains much better than the former. Further genetic analysis, combined with information describing the movement

of animals between herds, has tracked the spread of MAP strains between herds. Select variations which may impact the infectious ability of some strains were also identified.

Finally, to examine the occurrence of MSIs and microevolution within a selection of animals, multiple MAP samples from 14 cattle were examined using genomic sequencing. Twelve animals showed signs of both microevolution and MSIs, though the remaining two only showed evidence of microevolution. Variations found in repeat sequences found at select locations within the MAP genome indicated the potential production of altered proteins, which may have implications for the infectivity of this pathogen.

Acknowledgements

I would first like to acknowledge and extend a sincere thank you to my co-supervisors Dr. Kapil Tahlan and Dr. Nathalie Bissonnette for accepting me onto this project and for their continued patience, support, teachings, and guidance. I have learned much from the both of you over these four years, and your constant support has helped make the best of some difficult situations, such as the pandemic. I would also like to thank my supervisory committee members Dr. Lourdes Peña Castillo and Dr. Suzanne Dufour for the guidance that they have provided over the course of this project.

I would also like to acknowledge the School of Graduate Studies and Department of Biology at the Memorial University of Newfoundland for providing me the opportunity to complete my Ph.D., and for the financial support provided. I would especially like to thank Dr. Dawn Bignell for the continued shared use of both lab space and equipment. I would also like to thank the Agriculture and Agri-Food Canada (AAFC) team at the Sherbrooke Research and Development Center, specifically Dr. Séverine Ollier and Jean-Philippe Brousseau. Without their extensive culture work and cataloguing, this project would not be possible, and their mentorship was greatly appreciated during my visit to Sherbrooke. I would also like to thank Dr. Jean-Simon Brouard of AAFC for his development of the initial code used to visualize the qPCR and ELISA data, as discussed in Chapters IV and V.

I am also thankful for Dr. Isdore Chola Shamputa and Alex Goudreau of the University of New Brunswick for their work on the systematic review described in Chapter III, as well as Dr. Franck Biet, Dr. Cyril Conde, and Thierry Couchard of the

National Research Institute for Agriculture, Food and the Environment in France for their guidance for the work described in Chapter IV of this thesis. I would like to thank Dr. Brian Boyle of Laval University for his extensive work with us in the development of the amplicon sequencing method described in the Appendices. I would also like to thank the Centre d'expertise et de services Génome Québec for their sequencing services and support. I acknowledge that the advanced computing resources used in this research was provided by the Digital Research Alliance of Canada, the organization responsible for digital research infrastructure in Canada, and ACENET, the regional partner in Atlantic Canada. I would also like to thank John Reynolds, Librarian Assistant Professor, University of Miami Miller School of Medicine for peer reviewing the search using the Peer Review of Electronic Search Strategies (PRESS) guidelines described in Chapter III.

I would also like to express my gratitude to all my lab mates from the Tahlan, Bignell, and Lang labs for their constant support and friendship over the years, including Hannah Perry, Jody-Ann Clarke, Jordan Wight, Brandon Piercey, Arshad Shaikh, Gustavo Diaz Cruz, Madelyn Swackhamer, Menus Garg, Corrie Vincent, Kajal Gupta, Dr. Lancy Cheng, Dr. Phoebe Li, Dr. Jingyu Liu and Dr. Rupesh Sinha. Special thanks to Benjamin Constantine-Wall for working with me during his honours project.

Finally, last, and certainly not least, I would like to thank my parents and friends for their constant support both before and during this entire experience. I was only able to do this degree at all thanks to their continued support, both financially and mentally. As I move on to the next stage of my life, I'm glad I had you by my side throughout the entire experience. Thank you.

Co-authorship Statement

I am the primary author of the work conducted within the materials and methods described in Chapter II and the results and discussion sections discussed in Chapters III, IV, and V of this thesis.

Chapter III: The framework, development of key terms, primary screening through Covidence, data extraction and analysis, writing and primary editing were done by both me and I.C. Shamputa. A. Goudreau formulated the search strategy used to identify publications for the review. K. Tahlan helped develop the framework of the review, key term selection, conflict resolution on paper inclusion during Covidence analysis, and the primary writing and editing of the manuscript. N. Bissonnette provided additional technical and editorial input throughout the manuscript writing and editing process.

Byrne, A. S., Goudreau, A., Bissonnette, N., Shamputa, I. C., & Tahlan, K. (2020). Methods for Detecting Mycobacterial Mixed Strain Infections-A Systematic Review. Frontiers in Genetics, 11, 600692. https://doi.org/10.3389/fgene.2020.600692

Chapter IV: All genomic analysis conducted in this Chapter was performed by me. The laboratory of Dr. Bissonnette was in charge of the culturing of MAP strains. S. Ollier and other lab members collected MAP isolates from initial fecal samples (including mycobacterial culturing, MAP phenotype determination, cryopreservation, and axenic culture growth) and preparation for sequencing (DNA extraction). The collection of qPCR and ELISA data, along with the ANOVA and Tukey Analysis of SNPs **(Figure 4.3)**, was also performed by N. Bissonnette's research group. The initial draft of the manuscript was

performed by me, S. Ollier, and N. Bissonnette, with K. Tahlan and F. Biet contributing to further writing and editing. The design of the study was performed by N. Bissonnette, with both N. Bissonnette and K. Tahlan also responsible for project funding, administration, and supervision.

Byrne, A., Ollier, S., Tahlan, K., Biet, F., & Bissonnette, N. (2023). Genomic epidemiology of *Mycobacterium avium* subsp. *paratuberculosis* isolates from Canadian dairy herds provides evidence for multiple infection events. Frontiers in Genetics, 14. https://www.frontiersin.org/articles/10.3389/fgene.2023.1043598

Chapter V: All genomic analysis and protein prediction work conducted in this Chapter was performed by me. S. Ollier collected MAP isolates from initial fecal samples (including mycobacterial culturing, MAP phenotype determination, cryopreservation, and axenic culture growth) and preparation for sequencing (DNA extraction). N. Bissonnette provided the collection of qPCR and ELISA data visualized within this Chapter. The initial draft of the manuscript was performed by me and K. Tahlan, with S. Ollier and N. Bissonnette contributing to further writing and editing. The design of the study was performed by N. Bissonnette, with both N. Bissonnette and K. Tahlan also responsible for project funding, administration, and supervision.

This manuscript has been accepted for publication in the journal *Microbiology Spectrum*. **Byrne, A.**, Bissonnette, N., Ollier, S., Tahlan, K. Investigating *in vivo Mycobacterium avium* subsp. *paratuberculosis* microevolution and mixed strain infections.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

Definition	Abbreviation
6-carboxy fluorescein	6-FAM
Acquired Immunodeficiency Syndrome	AIDS
Agar Gel Immunodiffusion	AGID
Agriculture and Agri Food Canada	AAFC
Analysis of Variance	ANOVA
Average Nucleotide Identity	ANI
Bacille Calmette-Guérin	BCG
Base Pair	bp
Basic Local Alignment Search Tool	BLAST
Brain Heart Infusion	BHI
Clustered Regularly Interspaced Short Tandem Repeats	CRISPR
Colony-Forming Units	cfu
Complement Fixation Test	CFT
Crohn's Disease	CD
Cycle Threshold	Ct
Deoxynucleoside Triphosphate	dNTP
Deoxyribonucleic Acid	DNA
Dimethyl Sulfoxide	DMSO
Direct Repeat	DR
Discriminatory Index	DI
Enzyme-Linked Immunosorbent Assay	ELISA
European Nucleotide Archive	ENA
Genome Equivalent	Ge
Genome-to-Genome Distance	GGD
Green Fluorescence Protein	GFP
Hexadecylpyridium Chloride	HPC
Higashi Nagoya Tandem Repeats	HNTR
High-Performance Liquid Chromatography	HPLC
Human Immunodeficiency Virus	HIV
Insertion Sequence	IS

Interactive Tree of Life	ITOL
Internal Transcribed Spacer	ITS
Johne's Disease	JD
Large Sequence Polymorphism	LSP
Latin American & Mediterranean	LAM
Mammalian Cell Entry	mce
Microfold Cells	M-cells
Mixed Genotype Infection	MGI
Mixed Strain Infection	MSI
Multi-locus Short Sequence Repeat	ML-SSR
Multi-Locus Variant Analysis	MLVA
Mycobacteria Other Than Tuberculosis	MOTT
Mycobacterial Interspersed Repetitive Unit	MIRU
Mycobacterial Membrane Protein Large	mmpL
Mycobacterium avium Complex	MAC
Mycobacterium avium subsp. *paratuberculosis*	MAP
Mycobacterium avium Tandem Repeats	MATR
Mycobacterium tuberculosis complex	MTBC
National Center for Biotechnology Information	NCBI
Newfoundland and Labrador	NL
Non-tuberculous Mycobacteria	NTM
Nucleotide	nt
Oleic-Albumin-Dextrose-Catalase	OADC
Ontario	ON
Open Reading Frame	ORF
Peer Review of Electronic Search Strategies	PRESS
Phosphatidylethanolamine-Binding Protein	PEBP
Predicted Local Distance Difference Test	pLDDT
Preferred Reporting Items for Systematic Reviews and Meta-Analyses	PRISMA
Proline-Glutamate/Proline-Proline-Glutamate	PE/PPE
Pulse Field Gel Electrophoresis	PFGE
Quantitative Real-time Polymerase Chain Reaction	qPCR
Quebec	QC

Raf-Kinase Inhibitor Proteins	RKIP
Random Amplified Polymorphic DNA	RAPD
Restriction Fragment Length Polymorphism	RFLP
Ribosomal Deoxyribonucleic Acid	rDNA
Ribosomal Ribonucleic Acid	rRNA
Sample/Positive	S/P
Sequence Read Archive	SRA
Short Sequence Repeat	SSR
Signal Peptidase II	SPII
Single Nucleotide Polymorphism	SNP
Slipped Strand Mispairing	SSM
Spacer Oligotyping	Spoligotyping
Statistical Analysis System	SAS
The Centre for Applied Genomics	TCAG
Variable Number Tandem Repeat	VNTR
Whole Genome Sequencing	WGS

CHAPTER I

Introduction

1.1 Importance of the Dairy Industry

The Canadian dairy industry, which extends to both dairy farms and dairy production plants, is a significant contributor to the economy, with the total net farm cash receipts from dairying amounting to $8.23 billion in 2022 [1]. The close contact of a high number of animals within dairy farms makes animals more vulnerable to the spread of infectious diseases. Animal diseases caused by pathogens that spread throughout farms can be costly due to treatment expenses, a decrease in production or, in some cases, the need to cull the animal [2–4]. Alongside the immediate costs from either a strictly economic or animal health perspective, some diseases are zoonotic, having the potential to spread to humans. Such zoonotic pathogens involve additional costs in both the food safety and human health fields, requiring further investment into the development and execution of biosecurity measures and control programs for at-risk farms. One such potentially zoonotic pathogen which deserves special attention is the causative pathogen of Johne's Disease (JD), *Mycobacterium avium* subsp. *paratuberculosis* (MAP) [5–8]. The annual economic impact of MAP has been estimated at $28 million USD in Canadian farms [9], due to reduced milk production, increased cattle mortality, and early culling of infected animals.

1.2 The genus *Mycobacterium*

The genus *Mycobacterium* includes 195 different species with diverse growth characteristics and host tropism [10–12]. Mycobacteria can be categorized into three major

groups: those that cause tuberculosis (TB) and are part of the *Mycobacterium tuberculosis* complex (MTBC) [13], those that cause leprosy (including *Mycobacterium leprae* and *Mycobacterium lepromatosis*) [14], and all remaining members, collectively referred to as either atypical mycobacteria [15], non-tuberculous mycobacteria (NTM), or mycobacteria other than *M. tuberculosis* (MOTT) [10]. Additionally, members from this genus can be further categorized based on their growth rates into rapid and slow growers, with the latter having prolonged doubling times that make cultivation difficult [16–18]. Recent works using genomic analysis methods have noted the division of the larger *Mycobacterium* genus into five smaller clades, and have proposed a taxonomic reclassification which denotes each of these clades as separate genera [16,19]. In addition to the reclassified *Mycobacterium* genus, which includes many major human pathogens, the other clades were renamed as *Mycolicibacterium*, *Mycolicibacter*, *Mycolicibacillus*, and *Mycobacteroides* [16,19,20].

1.3 *Mycobacterium avium* subsp. *paratuberculosis* (MAP)

1.3.1 The *Mycobacterium avium* complex

Several species of non-tuberculous mycobacteria are naturally occurring opportunistic pathogens. They may inhabit several environments [21,22], including both natural and drinking water [23–25], soil [26–28], and dust or aerosols [29,30]. Among non-tuberculous mycobacteria, members of the slow-growing *Mycobacterium avium* complex (MAC) are the most frequent causative agents of disease in humans, able to cause pulmonary disease [26,31,32], lymphadenitis [33–35], and disseminated infections in severely immunocompromised hosts, such as those infected with human immunodeficiency virus

(HIV) [36,37] or those under treatment with immunosuppressants [38]. Initially, only the species *Mycobacterium avium* and *Mycobacterium intracellulare* were classified within the MAC, differentiated based on whether the bacteria were pathogenic to birds (*M. avium*) or humans (*M. intracellulare*) [39]. Through the use of modern molecular identification methods, several new species have been added to the MAC over time, namely *Mycobacterium intracellulare* subsp. *chimaera*, *Mycobacterium colombiense, Mycobacterium arosiense*, *Mycobacterium vulneris, Mycobacterium marseillense, Mycobacterium timonense, Mycobacterium bouchedurhonse, Mycobacterium mantenii, Mycobacterium intracellulare* subsp. *yongonense, Mycobacterium paraintracellulare,* and *Mycobacterium senriense* **(Table 1.1)** [40–48]. Differentiation of MAC species is typically done through examination of the high-performance liquid chromatography (HPLC) patterns of the mycobacterial cell wall components [39,49] or through multi-locus sequence analysis of select portions of MAC genomes: the 16S rRNA gene, *hsp65*, *rpoB*, and the 16S-23S rDNA internal transcribed spacer (ITS) [39–41,43,45–47].

1.3.2 *Mycobacterium avium* subspecies

All subspecies of *M. avium* are acid-fast, slow-growing bacterial pathogens which may cause infections in a variety of host species. *M. avium* subspecies are all closely related, sharing over 96% average nucleotide identity (ANI) and over 70% genome-to-genome distance (GGD) [50–52]. Comparing isolates to determine if they belong to the same species or subspecies is possible using select ANI or GGD thresholds. Pairs of taxa found to have < 95% ANI similarity (70% GGD) are different species, while those > 95% ANI similarity (70% GGD) are considered the same species [52]. Similarly, any taxa pairs with >

97% ANI similarity (80% GGD) are classified as the same subspecies [52]. In addition to ANI or GGD based analysis, a selection of typing and subtyping methods, allowing for both the identification of a specific MAC subspecies along with the identification of subspecies specific lineages, is possible. MAC subspecies typing is often performed through the analysis of select insertion elements, including *IS1245, IS1311, IS900,* and *IS901* [53–55], or based on the presence or absence of large sequence polymorphisms (LSPs) [56]. Confirmation of *M. avium* subsp. *lepraemurium* is possible through the confirmation of select polymorphisms found within the 16S rRNA gene [52,57].

M. avium subsp. *avium*, alongside non-MAC species *Mycobacterium genavense*, is the causative element of avian tuberculosis; a disease with a global prevalence that results in emaciation, diarrhea and formation of caseous tubercular nodules and granulomas in the liver, spleen, intestine and bone marrow of birds [58–60]. In addition to causing disease within both wild and domestic birds, *M. avium* subsp. *avium* can infect mammalian species, including humans [58,60–63].

M. avium subsp. *silvaticum*, formerly referred to as the "wood pigeon mycobacteria", has been known to infect a variety of species [64–67]. Due to the high sequence similarity between *M. avium* subsp. *silvaticum* and *M. avium* subsp. *avium*, and the availability of only a single type strain, recent studies have suggested that *M. avium* subsp. *silvaticum* should not be counted as a separate subspecies [52,57,64].

M. avium subsp. *hominissuis* is an opportunistic pathogen which favours infecting mammalian species such as humans [68], pigs [61] or dogs [69]. This is currently the most clinically relevant *M. avium* subspecies in immunocompromised patients, with patients

acquiring disseminated infections, soft tissue infections, pulmonary infections or cervical lymphadenitis [32,61,68,70–72].

M. avium subsp. *lepraemurium* (formerly *M. lepraemurium*) is the causative agent of both murine and feline leprosy. This subspecies was initially thought to be closely related to the human pathogen *M. leprae* due to both similar disease symptoms and the difficulty of growing both species in axenic culture [73–76]. However, recent phylogenomic work has shown that this species is a member of the MAC, and has been incorporated as a subspecies of *M. avium* [52,57,77].

Mycobacterium avium subsp. *paratuberculosis* (MAP) is a pathogenic organism which is the causative element of JD, a chronic granulomatous enteritis which primarily affects both wild and domestic ruminants [78–80]. MAP has also been associated with several human diseases, the most notable of which is Crohn's disease (CD). CD is both clinically and histopathologically similar to JD, and while this suggests that mycobacterial infection may have an impact on CD, the exact role of MAP in the development of CD remains controversial due to the lack of definitive evidence [81–84]. MAP has also been associated with several other human diseases, including Blau syndrome [85], type I diabetes mellitus [86–91], Hashimoto's thyroiditis [92–94] and multiple sclerosis [95–102].

MAP is closely related to *M. avium* subsp. *avium*, with a 16S rRNA sequence similarity of 99.9% [103]. Like other members of the MAC, MAP is a slow-growing mycobacteria, with a doubling time exceeding 22-26 hours [17,104]. MAP has some features that make it distinct from other members of the MAC and other subspecies of *M. avium*. One difference is that most strains of MAP require the compound mycobactin to be

present within media to grow under laboratory conditions, unlike other members of the MAC [66,105–107]. Mycobactin is a siderophore, an iron-chelating compound that assists in the transport of iron across the cell membrane that is typically produced by members of the *Mycobacterium* genus to allow for enhanced survivability within macrophage phagosomes [108–110]. Due to a truncation within the *mtbA* gene of the mycobactin cluster, MAP is unable to produce mycobactin [111,112], instead potentially using an alternate iron uptake pathway to compensate for this loss [113,114].

Genomic features unique to MAP include the presence of insertion sequences *IS900*, a 1.4 kb multicopy (16-22) insertion sequence (IS) exclusive to the MAP genome [57,115,116], and *ISMap02*, a multicopy (6) putative IS element found only within the MAP genome [117,118]. *IS900* has been used for strain typing of MAP strains through restriction fragment length polymorphism (RFLP) analysis [119]. However, typing by RFLP using IS900 is limited, as some studies have indicated that PCR primers traditionally used for *IS900* region amplification can amplify *IS900*-like sequences present in other mycobacterial species [120,121]. Such studies have found that non-MAP isolates had *IS900*-like sequences with 71% to 94% homology to *IS900.* Recent work has shown that primers previously used as the standard for *IS900* typing have mismatches which cause amplification of these *IS900*-like sites, and recommend the use of updated primer pairs that are more specific to *IS900* [122,123].

Similarly to *IS900*, the multicopy and MAP-exclusive nature of *ISMap02* made it a target for PCR-based MAP-detection assays of fecal samples [124,125]. However, like *IS900*, it was found that the PCR showed positive results with non-MAP samples,

suggesting a conserved region between *ISMap02* and non-MAP IS elements [126]. Aside from *IS900*, several single-copy genes unique to MAP have been identified [127] and some have been used for detection purposes including *hspX* [128–130], *F57* [131–133], and locus 251 [134–136].

1.4 Johne's Disease

JD is a chronic granulomatous enteritis which is commonly found in domesticated ruminants such as cattle or sheep [137–145]. Detection of MAP has also been recorded in wild ruminants [142,146–150] and a variety of non-ruminant species including birds, foxes and non-human primates [151–155]. The development of JD is a complex process, with MAP incubating for a period of two to ten years after the initial infection of cattle before the development of clinical signs [156–158]. Shedding of MAP through feces is unpredictable and can occur within weeks of infection, contaminating the environment and spreading MAP throughout the herd [137,159,160]. Transmission between animals typically occurs through the fecal-oral route, with young cattle becoming infected by adult stock [161,162]. This is especially prevalent in calves younger than six months old, which are more susceptible to infection [163]. Transmission of MAP from cow to calf has also been recorded through the direct ingestion of colostrum or milk [164,165], though *in-utero* infections [166–169]. Infections derived from the environment have also been described [170,171].

1.4.1 Symptoms of Johne's Disease

JD is observed within cattle as part of a four-stage infection process, which is categorized based on the fecal shedding of MAP into the environment, the clinical

symptoms that are observed, and the ease at which MAP can be detected within infected organisms **(Figure 1.1)** [156,157,172].

The first stage of infection is referred to as the "silent infection" phase and occurs in calves up to two years old [170,172]. Contraction of MAP and progression to this phase typically occurs through the ingestion of contaminated milk, water or feed [164,165,173], as previously described. Cattle in the silent infection phase are asymptomatic, showing neither the sub-clinical nor clinical symptoms associated with later stages of JD, with no cost-effective manner to detect MAP at this stage [172]. Only direct post-mortem tissue culturing of MAP, which may take a significant amount of time, or histological detection methods can confirm MAP infection during the silent phase [172].

Animals in the second stage of infection, referred to as "subclinical infection" [172], show no clinical symptoms of infection at this phase, though reduced milk production and fertility has been reported [174]. However, animals at this stage may have developed antibodies against MAP, along with altered cellular immune responses typically associated with disease [172,175,176]. Detection through fecal cultures remains possible at this stage but, like fecal PCR, it is not sensitive due to the intermittent excretion of MAP [159,177,178]. Due to this intermittent shedding, MAP-positive cows with negative test results may continue to shed MAP into the environment, potentially allowing for the continued spread of MAP throughout the herd [157,162]. While once considered to be the gold standard for MAP confirmation, factors of fecal culturing such as the limited sensitivity, and the costly, time consuming, labour intensive nature of culturing MAP limits the efficiency of MAP detection [179,180]. MAP detection is also possible using fecal qPCR based methods,

which vary in efficiency depending on the commercial qPCR and DNA extraction kit used [178]. Enzyme-linked immunosorbent assay (ELISA) antibody testing (though results for the latter are often low at subclinical stages), may also be used to confirm infections [175,178].

The third stage of infection, "clinical infection", occurs when clinical symptoms emerge, and cattle health begins to noticeably deteriorate [157,172,181,182]. Clinical symptoms of JD may not emerge until two to ten years of incubation, with initial symptoms including weight loss and a potentially increased appetite. Several weeks after these initial symptoms, intermittent periods of diarrhea become visible, with additional symptoms such as increased thirst and lowered milk production also becoming prominent. At this stage of JD, fecal culture-based tests should reliably test positive, as will commercial assays such as ELISA and agar gel immunodiffusion (AGID) tests [172,183]. Due to the lowered milk production and deteriorating health conditions of the animal, along with any potentially positive JD results that are noticed in the regular screening of cow herds, infected cattle are often culled from the herd at the clinical stage [184].

If the infected animal is not culled at this point, a fourth stage of JD, referred to as "advanced clinical infection", will develop [157,172,185]. Animals at this extended stage of infection will become weak, lethargic, and emaciated. Diarrhea also changes from an intermittent state to a more consistent fluid-like state referred to as "pipe stream" diarrhea and additional symptoms, such as the development of intermandibular edema (bottle jaw) due to hyperproteinemia, cachexia and anemia. After the development of such symptoms, the cow will succumb to its infection.

1.4.2 Johne's Disease Mechanism of Infection

Upon ingestion of MAP, primary cell invasion occurs within the small intestine, specifically within the jejunal and ileal regions [107,186,187]. MAP primarily targets the microfold cells (M-cells) found within the Peyer's patches. These are small regions of organized lymphoid tissue that respond to the presence of pathogens within the small intestine [188], though uptake through differentiated epithelial cells has also been recorded [189,190]. M-cells act as invasion points for several pathogens aside from MAP, including enteropathogenic *Escherichia coli*, *Vibrio cholerae*, *Shigella flexini* and some *Salmonella* and *Yersinia* species [191,192]. Once in the small intestine, MAP expresses a series of fibronectin attachment proteins which promote bacterial attachment to fibronectin, and the subsequent opsonization of MAP [187,193,194]. This fibronectin-dependant uptake of MAP may contribute to the favouring of invasion of MAP through M-cells, which are enriched with the β1 integrins that act as fibronectin receptors [195]. Invasion of the M-cells occurs quickly, with cells able to pass through the barrier within 30 minutes [196].

Once MAP cells have crossed the epithelial layer of the intestinal mucosa, they are translocated into the submucosa, where they encounter dendritic cells [189,197–199] and subepithelial macrophages [186,200]. Uptake of MAP by macrophages may involve one or more families of receptors which are associated with mycobacterial entry, with different routes of entry having different impacts on intracellular survival, cytokine secretion and immune response [186,201]. An essential factor in the survivability of pathogenic mycobacterial species within the host is the manipulation of several macrophage functions, such as phagolysosome formation, macrophage responsiveness, and apoptosis

of the macrophage, to enhance immune evasion and allow for growth [186,187,202–205]. Once the intracellular environment has been adjusted to allow for MAP growth, it undergoes the long asymptomatic period associated with the silent infection stage, with some fecal shedding possible [206]. Once within the macrophage, granulomas begin to develop and grow within the host, progressing into complex, structured lesions as the infection continues [207,208]. Granulomas begin as focal granulomas, small granulomas relegated to intestinal lymphoid tissue or lymph nodes, before progressing as multifocal granulomas, where several small granulomas develop beyond the Peyer's patches, typically in lymphoid tissue or the lamina propria [207–209]. Cattle showing clinical symptoms of MAP typically show diffuse lesions, with widespread granulomatous enteritis and a thickening of the cell wall [207–209]. Diffuse lesions are described as being either paucibacillary, intermediate, or multibacillary, and have been shown in cattle, sheep, goats, and deer [207,210–212]. Lesions may be classified as paucibacillary when the enteric inflammatory infiltrate is primarily composed of lymphocytes with few mycobacteria [207–209]. Intermediate lesions develop when most of the inflammatory cells are giant cells or when macrophages contain only a few mycobacteria [207–209]. Multibacillary lesions occur when most inflammatory cells of the lesion are macrophages filled with numerous mycobacteria [207–209].

1.4.3 Diagnosis of MAP infection

Diagnostic testing to confirm that an animal or herd is free from MAP infection may be performed to allow for the trade of individual animals, estimate the prevalence of infection for use in surveillance or eradication programs, and confirm clinical cases of JD

[213–215]. Test accuracies may vary depending on how valid certain tests are for specific purposes, with tests broadly categorized depending on if the test is detecting the agent of infection (MAP) or detecting the host immune response [215].

Post-mortem necropsy with culture and histopathology on multiple tissues throughout the digestive tract allows for definitive confirmation of JD infection [215]. Histopathological examination (Ziehl-Neelsen staining) of tissue samples, such as the mesenteric and ileocecal lymph nodes, ileum, and liver [215,216], or fecal smears [217] allow for the detection of acid-fast bacteria within either lesions present throughout the tissue or fecal samples from shedding cattle. Confirmation of JD using culture-based methods is considered to have 100% specificity [215,218,219], allowing for the direct growth and subsequent confirmation of MAP from fecal or tissue samples. However, culturing MAP is labour intensive, as the fastidious nature of MAP causes cultures to take a long time to grow. In addition to this slow growth, culturing of fecal samples has a low sensitivity, as infection in animals with intermediate of low amounts of MAP being shed in the feces are not as easily detected using culture based methods [215,218,219]. DNA probes targeting MAP specific sequences allow for the rapid identification of bacterial isolates from a number of different sample types, including blood, milk, feces, tissues or environmental samples [215]. As previously discussed, analysis methods which rely on targeting MAP specific sequences include *IS900*-RFLP [119], and PCR targeting either *hspX* [128–130], *F57* [131–133], or locus 251 [134–136]. Real time qPCR using MAP specific targets also allows for the rapid detection of MAP, though the efficacy of the method is dependant on the quality of the DNA produced by the DNA extraction method used [178].

As with fecal PCR based detection methods, the sensitivity of immune based detection methods is generally low. This is especially the case in earlier stages of MAP infection, as in these early stages, the host will show a strong cell-mediated response, with the humoral response typically activating as the infection progresses [209,218,220]. ELISA tests for MAP are the most commonly used of currently available detection methods, being widely commercially available, rapid, cheap, and very specific, able to detect anti-MAP antibodies in both serum and milk samples [184,221]. While both agar gel immunodiffusion (AGID) and complement fixation tests (CFT) also act as measures of humoral response, they are less commonly used than ELISA test and have lower sensitivity [215,222,223].

1.5 Survival of MAP within the Environment

While MAP's role as an obligate intracellular pathogen limits growth to the presence of a host organism, MAP can persist within the environment for an extended period of time in both soil and water. Examination of MAP within dam water and sediment showed survival of MAP for 48 weeks when under shade and 36 weeks when somewhat exposed, with survival in sediment being 12 to 26 weeks longer than survival within the water column [224]. Similar survival experiments examining the persistence of MAP on soil and grass observed survival for up to 55 weeks in a fully shaded environment and only two weeks in environments without shade or covering vegetation [225]. MAP moves slowly through soils and tends to remain on the grass and upper layers of pasture soil, posing a risk of infection to grazing animals if a pasture is contaminated directly by an infected animal or indirectly through rainfall or irrigation [226,227].

While the exact mechanisms that allow for persistence outside of the host are not certain, one hypothesis suggests that the ability of MAP to survive ingestion by environmental protozoa may contribute to bacterial longevity [228,229]. Prior work examining the interactions of *M. avium* with the water-borne free-living amoebae *Acanthamoeba castellanii* has shown that *M. avium* was able to survive within the protozoan, able to inhibit lysosomal fusion and replicate similarly to growth observed within macrophages [230], while also being protected from the effects of select antimicrobials [231]. Additionally, the growth of *M. avium* within the amoeba was shown to be more virulent within mouse models than broth-grown bacteria [230]. Similar experiments repeated with MAP showed that ingestion and growth within *A. castellanii* [229,232] or *Acanthamoeba polyphaga* [233], a protozoa species commonly found in the same soil environment as MAP [234], was also possible. It was also found that MAP ingested by amoebae were more resistant to inactivation by chlorine than free MAP cells, suggesting that ingestion by amoebae may enhance MAP survivability in chlorinated surface water [229]. Other work confirmed that both Cattle-type (C-type) and Sheep-type (S-type) MAP strains were able to be ingested by *A. castellanii*, [228].

Additional hypotheses which may allow for the environmental persistence of MAP include the use of biofilms or the development of a spore-like morphotype. In certain circumstances, bacteria may form aggregates where cells are embedded within a matrix of extracellular polymeric substances, referred to as a biofilm [235]. Biofilm formation has been described in *Mycobacterium smegmatis, M. tuberculosis*, and *M. avium* subsp. *avium* [236–239]. The formation of glycopeptidolipids, species-specific

mycobacterial lipids found within the cell envelope [240,241], is an important factor in biofilm formation within both MAP and *M. avium* subsp. *avium* [242,243]. While the exact mechanism for biofilm formation in MAP is uncertain, transposon insertions within the non-ribosomal peptide synthase gene *pstA* resulted in reduced biofilm formation with loss of a specific lipotripeptide, suggesting the involvement of this gene in biofilm formation [243]. In the environment, MAP is commonly found within mixed-species biofilms within pipes, faucets and water troughs, allowing for spread to both humans and animals alike [244–246]. In addition to biofilm formation, MAP has also been reported to develop a spore-like form when placed under nutrient-starving conditions [247]. While entry into such a state would allow for MAP to persist under harsh conditions within both aquatic and soil environments, the ability of mycobacteria to sporulate is a controversial topic [247,248] and no further work has been performed to examine the impact of these spore-like morphologies on MAP infections.

1.6 Control of Johne's Disease

Protection of a herd against JD is a unique challenge. The combination of limited methods of early detection, the ability of MAP to spread throughout the herd through fecal shedding, and the observation of disease symptoms at a much later time than when the initial infection occurs or the infectious phase begins make any treatment ability a complicated matter. Initial vaccines for MAP were ineffective, with whole-cell-based vaccines that were either live attenuated or inactivated showing that while clinical symptoms and colonization of MAP can be limited, the actual infection is not prevented [249]. Additionally, whole-cell MAP vaccines are also not favoured, as their use for

preventing MAP infection causes interference with the diagnosis of both bovine tuberculosis and legitimate paratuberculosis infections, making already difficult detection even more ambiguous [250–252]. Current vaccine developments for JD focus on the use of live attenuated vaccines [253–255] or subunits, such as recombinant MAP proteins with adjuvants [249,256–258]. An ideal vaccine for MAP should be able to limit interference with the testing of bovine tuberculosis and paratuberculosis, prevent the organism from reaching the clinical stage of infection, limit premature culling, and be able to produce anti-MAP antibodies [249].

Due to the limited options available for vaccines, a number of control programs have been developed in several countries to try and minimize the spread of MAP throughout herd populations [184,259]. The goal and strategies applied through control methods may change depending on factors such as the size of the herd, whether complete eradication is possible, and how new cattle are introduced to the herd [184]. Several control strategies use a combination of best management practices to limit the risk of cattle contracting MAP, in addition to a "test-and-cull" method that allows for the frequent examination of herds, the rapid removal of infected cattle, and the allowance of farmers to adjust to the loss [260,261].

Due to its nature as an obligate intracellular pathogen, MAP requires host cells to survive and multiply, and the elimination of MAP-infected animals from the herd should theoretically lead to the eradication of disease within the herd [262]. This idea is the foundation on which many JD control programs are based, and while some cases have shown a significant reduction in the prevalence of JD [263–265], complete eradication is only

achievable in some herds [266,267]. The ability for MAP infections to exist within a host during the silent and sub-clinical stages before symptoms appear, the shedding of MAP into the environment, and the low sensitivity of available tests are all factors that lead to the enhanced spread of undetected MAP throughout the environment [157,162]. While animals showing obvious signs of infection can be removed and culled, there may still be several animals with MAP that remain undetected by conventional testing and continue to spread it into the environment, referred to as the "iceberg effect" [156,172]. Initial estimates proposed that for every cattle showing advanced clinical symptoms, up to 14 cattle were within the silent stage of disease and that more cattle were in the silent infection stage compared to the subclinical and clinical stages [172]. However, more recent models suggest that there are fewer silent infections than previously suggested and that the majority of difficult-to-detect cattle are instead in the subclinical stage of infection **(Figure 1.1)** [156].

1.7 Sub-typing and Strain Typing of MAP

While on-farm biosecurity measures aim to limit the spread of infection [268], the introduction of novel cattle into a herd presents a risk of transmission. Due to the slow-growing, persisting nature of JD, it is important to establish the source of infection as quickly as possible. Molecular characterization methods are proving to be powerful tools, able to track infections, and aid in forming legislation to enhance control methods, and limit the spread of JD [269–271]. MAP strains can be broadly characterized into sheep-type (S-type/Type I and III) and cattle-type (C-type/Type II) sub-types, with the latter also comprising of the bison-type (B-type) sub-type. Differentiation between these subtypes can be determined based on the presence or absence of select large sequence

polymorphisms (LSP) [272,273] and differences within the *IS1311* insertion sequence [274–276]. The large sequence polymorphism LSP^{A}20 is absent in S-type strains, while LSP^{A}4 is absent in C-type strains [272,273]. The confirmation of these LSP sequences through PCR allows for accurate sub-typing. Restriction enzyme assays targeting the *IS1311* insertion sequence allow for the differentiation between S-Type and C-Type sub-types based on the presence of a *HinfI* cleavage site caused by a C/T SNP within select copies of the insertion sequence [277]. The presence of this SNP within a C-type *IS1311* PCR product allows for the visualization of a distinct *HinfI* restriction enzyme digestion profile after size separation by agarose gel electrophoresis [277]. Additional works show that the division into two major sub-types can also be determined using *IS900*-RFLP or Pulse Field Gel Electrophoresis (PFGE) based methods, which are now referred to as Type I and Type II [278,279].

MAP isolates which would later be classified as part of the Type III sub-type were initially classified as an intermediate strain between Type I and II, based on hybridization patterns observed in restriction enzyme analysis with DNA probes, *IS900*-RFLP, and PFGE methods [274,275]. Whole genome sequencing (WGS) results later confirmed that Type III isolates were a sub-type of Type I [53]. Distinction of B-type isolates from Type II isolates was initially performed using *IS1311* PCR restriction endonuclease assays [276,280], with both *IS1311* PCR restriction endonuclease assays and WGS later confirming the presence of "American Bison" and "Indian Bison" lineages [281,282].

It has been noted that MAP isolates from different lineages are not specific to a particular host; therefore, the Type I, II and III classification system is preferred [53,271,283–

[285]. Over time, the examination of the MAP genome has also allowed for the development of strain typing methods based on analyzing the differences in the copy numbers or sequences of variable DNA repeats. These include the eight locus Mycobacterial Interspersed Repetitive Unit-Variable Number Tandem Repeat (MIRU-VNTR), and the eleven locus Short Sequence Repeat (SSR) typing methods, both of which are commonly used in epidemiology and source tracking studies [286,287]. More recently, whole genome sequencing (WGS) followed by analysis of single nucleotide polymorphisms (SNPs) has enabled strain typing at a much higher resolution than previously possible, leading to a greater understanding of the MAP genome[53,288–290].

1.8 Mixed Strain Infection (MSI)

The progression and outcome of an infection is dependent on many factors, which include the resident-host microbiome and the presence of other pathogens, sometimes from the same genus **(Figure 1.2)** [291–293]. Mixed-species infections refer to the phenomenon where different species belonging to the same genus concurrently infect a single host. Another important factor to consider is the potential for genetically distinct strains (or isolates) of the same pathogenic species to infect a single host at any given time, referred to as a polyclonal infection [294–296]. Polyclonal infections can potentially arise if an isolate undergoes intra-host evolution (referred to as microevolution) following infection, leading to minor genetic differences in the resulting progeny [297–299]. Another mechanism leading to polyclonal infections involves concomitant or sequential infection by genetically distinct strains, a process referred to as mixed strain infection (MSI). The presence of genetically distinct variants of a single pathogen species in an infected host,

whether it be due to MSI or microevolution, is collectively referred to as a mixed genotype infection (MGI).

The presence of multiple strains with varying genotypes can result in altered physiological characteristics or pathogenicity, which in turn can affect transmission dynamics [296], or lead to treatment complications due to varying antibiotic resistance profiles (also known as heteroresistance) [300,301]. In these cases, a single treatment regimen may not be optimal against all strains in an individual, causing the infection to persist or return after briefly subsiding. MSIs are particularly relevant in slow-growing pathogens such as *M. tuberculosis*, MAC members and other related mycobacteria, as these organisms can remain undetectable for long periods of time [172,302]. If an MSI exists and the initial treatment is unsuccessful, the persistence of these infections may result in the development of more severe disease over time [303]. Additionally, MSIs have the potential to interfere with host immune responses due to antigenic variations that might exist between different strains [294,304,305].

1.9 Objectives of Thesis Research

The presence of MSIs influences both the progression, outcome, and transmission dynamics of tuberculosis infection in humans due to differences within the physiological characteristics of individual isolates [306–310]**.** In tuberculosis, treatment complications may be caused by heteroresistance in MGIs [300,301] . While antibiotic resistance in MAP has been identified in certain studies [311], antibiotics have not typically performed to clear MAP infections of JD, though some treatments for Crohn's Disease have been attempted [81,312]. This lack of approved antibiotic treatments against JD suggests that the

evolutionary pressure on MAP is not associated with antibiotic exposure, instead arising from prolonged infection of the host. The genome evolution rate of MAP is low, with estimates ranging from 0.1 to 0.5 SNPs/genome/year [53,290,313]. Due to the fact that cattle become infected with MAP at ages younger than six months [163] with very long periods of incubation (5-7 years) caused by this slow-growing pathogen we hypothesize that MGIs, and especially MSIs, could have implications during infections caused by slow-growing pathogens like MAP. Detection of blood immune markers or other biomarkers such as MAP excretion in stools used to screen for JD is unpredictable in subclinical animals, a long period of 5 to 7 years during which the animal spends most of its life. In other diseases, MGIs are known to interfere with the host immune response, possibly due to the different antigenic responses induced by the infecting strains [294,304,305]. Since mixed-strain tuberculosis infections impact the treatment and control of tuberculosis [294,314], it is hypothesized that MGI may also influence the manifestation of biomarkers in cows with JD. The presence of MSIs was previously confirmed for MAP within a single animal when combined with fragment analysis technology [315], which was later validated through WGS and SNP-based analysis [289]. Despite this, there is still ambiguity as to the nature of MSIs in MAP, especially in their identification, prevalence, and dissemination potential which this thesis will seek to address.

In Chapter III of this thesis, a systematic review was conducted to gain a better understanding of the prevalence, species, and location of MSIs within the genus *Mycobacterium* found within the available literature and the methods used to detect them. Additionally, this Chapter intended to clarify the difference between true MSIs as

compared to similar events such as re-infection, relapses, polyclonal evolution, and microevolution, as the exact definitions vary between studies [316–320].

Chapter IV of this thesis examines genetic variations between 67 MAP isolates grown from subclinical cows shedding the bacterium in their feces. These animals were from dairy herds located in Quebec (QC) and Ontario (ON), two major dairy-producing provinces in Canada. The use of WGS-based methods allowed for the examination of these strains through ML-SSR, MIRU-VNTR and SNP-based comparative methods, and allowed for comparison of their discriminatory capabilities and identification of multiple unrelated strains infecting cattle at the herd level. This WGS data, combined with animal health and movement data, also allowed for the tracing of infections within individual herds.

Chapter V of this thesis used in-depth investigative methods to examine the presence of MAP MGIs within fourteen individuals from three dairy cattle herds. WGS analysis was used to examine genetic differences in multiple MAP isolates derived from individual animals shedding high levels of the bacterium. Examination of ML-SSR, MIRU-VNTR and SNP patterns within isolates both within the same animal and between different animals allowed for the identification of MGIs, and in some cases, provided evidence for either microevolution or MSI events.

Additional unpublished work is also described within **Appendix III**, which details the attempted development of PCR-based assays intended for cheap, reliable, and reproducible sequencing of SSR homopolymer repeats beyond 15-nucleotides in length.

Both a fragment analysis-based approach using four SSR loci (1, 2, 8 and 9), alongside an amplicon sequencing-based method targeting SSR1 were attempted.

1.6 Figures and Tables

1.6.1 Figures

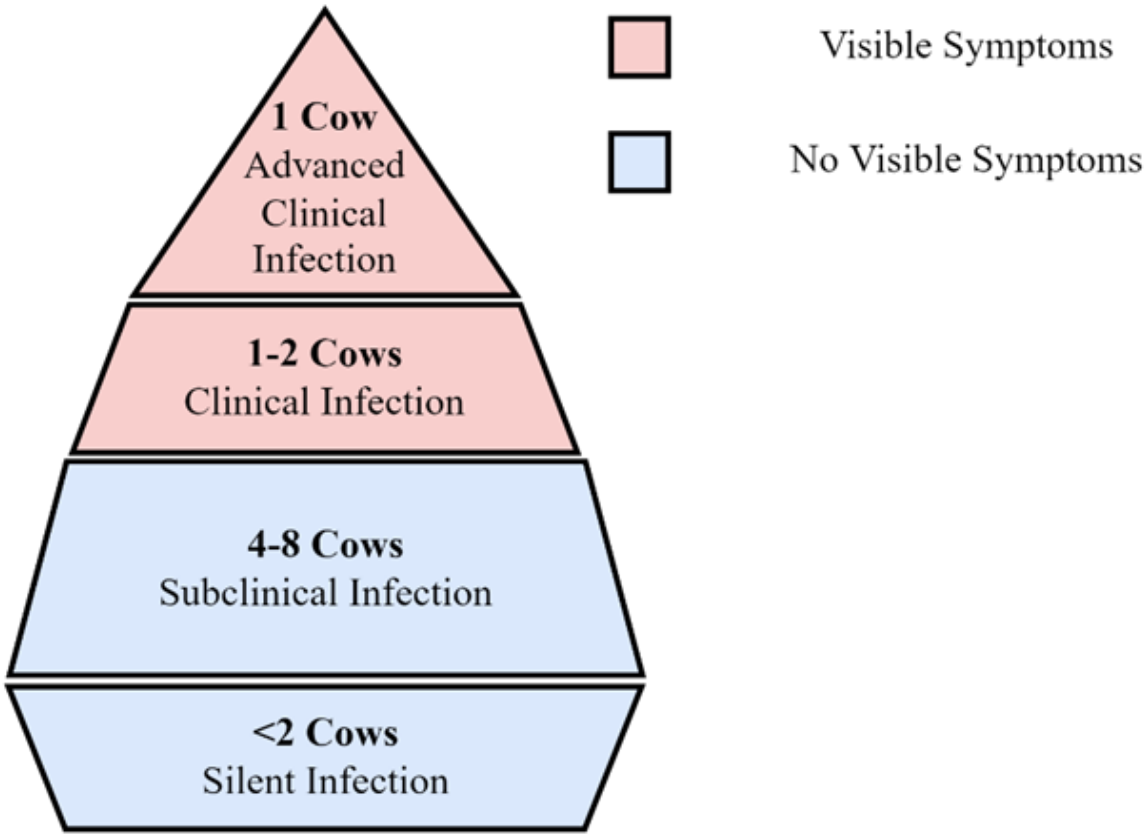


Figure 1.1: Stages of Johne's Disease. Stages are coloured based on whether symptoms of each stage are observable (red) or not observable (blue). The number of cattle at each stage represents the "iceberg effect" model described by Magombedze et al.[156], where for each cow within the advanced clinical stage, it is predicted that a select number of cattle are within earlier stages of Johne's Disease, even if no symptoms can be observed.

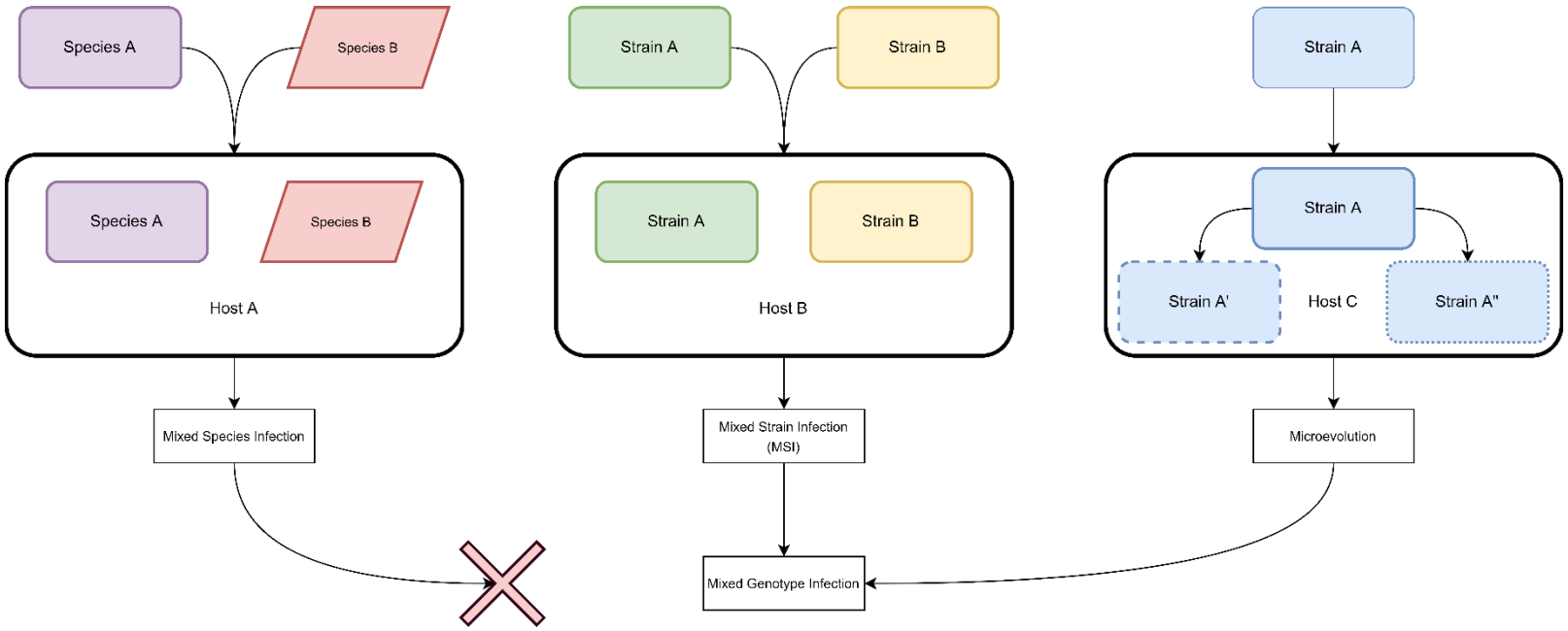


Figure 1.2: **Schematic distinguishing between mixed species infections, mixed strain infections (MSI), and microevolution.** Mixed species infections (left) describe instances where multiple species of pathogens infect a host (white rectangle) at the same time. During MSI, different strains of the same species infect a single host at the same time (middle). In microevolution, a single strain infects a host and undergoes within-host evolution, resulting in genetically related lineages (right). Both MSI and microevolution, but not mixed species infections, can be categorized under the umbrella term of mixed genotype infection (MGI).

1.6.2 Tables

Table 1.1: Species contained within the *Mycobacterium avium* complex (MAC) and their associated diseases

Species	Initial species description	Observed Infections
Mycobacterium aroseiense	[41]	Osteomyelitis and pulmonary infections in humans [321,322].
Mycobacterium avium subsp. *avium*	Establishment of subspecies by [66].	Primarily avian tuberculosis, some mammalian and human infections were noted [58,60–63].
Mycobacterium avium subsp. *hominissuis*	Establishment of subspecies by [323].	Soft tissue infections, pulmonary infections and cervical lymphadenitis in humans and pigs [32,61,68,70–72].
Mycobacterium avium subsp. *lepraemurium*	Initially described by [76], added to MAC by [77], declared *Mycobacterium avium* subspecies by [52].	Murine and feline leprosy [73–75].
Mycobacterium avium subsp. *paratuberculosis*	Establishment of subspecies by [66].	Johne's Disease in ruminant organisms [157], potentially Crohn's Disease in humans [82,83].
Mycobacterium avium subsp. *silvaticum*	Establishment of subspecies by [66].	Opportunistic infections in a variety of bird and mammal species [64,65,67].
Mycobacterium bouchedurhonense	[42]	Pulmonary Infections [42]

Mycobacterium colombiense	[324] as MAC-X, [45] as *Mycobacterium colombiense.*	Lymphadenopathy and pulmonary infections within humans with weakened immune systems, with fatal cases previously recorded [325–327].
Mycobacterium intracellulare subsp. *chimaera*	[324] as MAC-A, [46] as *Mycobacterium chimaera*, later amended to be a subspecies of *Mycobacterium intracellulare* by [20].	Pneumonia, opportunistic endocarditis after open heart surgeries [328–332].
Mycobacterium intracellulare subsp. *intracellulare*	Establishment of subspecies by [333].	Pulmonary infections [334,335].
Mycobacterium intracellulare subsp. *yongonense*	Misidentified as a species by [43], later amended to be a subspecies of *Mycobacterium intracellulare* by [333].	Pulmonary infections [336,337].
Mycobacterium mantenii	[48]	Skin and soft tissue infections [48,338–340].
Mycobacterium marseillense	[42]	Pulmonary infections, skin infections and lymphadenitis [341–345].
Mycobacterium paraintracellulare	[44]	Pulmonary infections [44].

Mycobacterium senriense	[40]	Pulmonary infections [40].
Mycobacterium timonense	[42]	Pulmonary infections, opportunistic infections [346,347].
Mycobacterium vulneris	[324] as MAC-Q, [47] as *Mycobacterium vulneris*.	Suppurative wounds, cervical lymphadenitis [47].

CHAPTER II

Materials and Methods

2.1 Systematic Review Procedure

A systematic review of the literature was performed to identify methods used for detecting MSIs caused by a single species of mycobacteria. This process was comprised of four stages, with the first stage entailing the use of a comprehensive search strategy to locate published studies. An initial limited search was conducted in Ovid MEDLINE to compile a list of keywords and index terms from relevant articles. A full search strategy was then developed by a librarian who tested search terms in MEDLINE and only those terms yielding unique results were retained for further use. The full search strategy was externally peer-reviewed by a second librarian using the Peer Review of Electronic Search Strategies (PRESS) guidelines [348]. The search strategy was then adapted for EMBASE (Elsevier) following which both the MEDLINE and EMBASE were initially searched on June 11, 2020, with no prerequisite limits **(Table 2.1 A** and **C)**. An updated search was performed on August 25th, 2020, to include the term "polyclonal" **(Table 2.1 B** and **D)** Search results were collated and uploaded into Endnote version X8.2 (Clarivate Analytics, PA, USA) for organization followed by Covidence™ (Veritas Health Innovation Ltd., Melbourne, Australia), a screening and extraction tool for systematic reviews. After the removal of duplicates, reference lists of all selected studies were screened for additional articles of interest using Google Scholar **(Figure 2.1)**. The second stage involved the screening of article titles and abstracts by two reviewers. The inclusion criteria for articles at this stage were as follows: (i) presence of both a title and abstract;

(ii) abstracts in English or French; (iii) primary article or review; (iv) explicit mention of mixed infection or other synonymous terms like double infection, multiple infection, simultaneous infection or polyclonal infection; (v) described molecular or phenotypic methods; and (vi) mention of *M. tuberculosis* (MTBC), non-tuberculous mycobacteria (NTM), mycobacteria other than tuberculosis (MOTT), atypical mycobacteria, environmental mycobacteria, or *M. leprae*. The presence of additional terms including "mixed, simultaneous, multiple, mycobacteria, concomitant, concurrent, co-infection, polyclonal, *Mycobacterium*, heterogeneity, subtype, sub-type, double, more than one, dual, and superinfection" were searched for within abstracts to assist in study identification. Articles were excluded if they were: (i) opinion pieces; (ii) commentaries; or (iii) not written in English or French. Conflicts on the inclusion/exclusion of specific articles were resolved after analysis by a third reviewer or by mutual agreement by the team.

Full texts of articles were uploaded to Covidence™ and subsequently assessed for eligibility at the third screening stage using the following inclusion criteria: (i) full-text availability; (ii) article in English or French; and (iii) reported mixed infection, double infection, multiple infection, simultaneous infection or polyclonal infections involving one *Mycobacterium* species using molecular or phenotypic methods. Articles were excluded from the main list if: (i) they reported clonal variants suggestive of microevolution of a single strain; (ii) reported detected mixed-species infections, e.g., *M. tuberculosis* and *M. bovis* together; (iii) the full text of the article was unavailable; (iv) they were dissertations, conference, or poster presentation abstracts; (v) did not contain

sufficient information; or (vi) were not in English or French. In the last stage of the analysis, initial findings from the full-text screening were analyzed to identify data for extraction and use in the review. All articles citing the identified reports for use in the review were also examined for relevant information, which was included in the final results.

2.2 Sample Selection, Growth of MAP, and Sequencing of Isolates

All reagents described in the following sections were purchased from Thermo-Fisher Scientific or MilliporeSigma, unless otherwise stated.

2.2.1 Animal Selection and Sample Isolation

The source population of cattle for work described in Chapters IV and V was selected from 22 Canadian herds based on the reported incidence of JD in cows during the years 2013-2017. All animal procedures described in Chapters IV and V obtained ethical approval from the Agriculture and Agri-Food Canada Animal Ethics Committee (AAFC-545).

Fecal samples were collected using a single-use veterinary glove. A blood sample was drawn from each JD-positive cow using an 8.5 mL Serum Separation Tube (Becton, Dickinson and Company, Franklin Lakes, NJ, SKU:367988). Blood tubes were centrifuged at 1,200 x g at 4°C for 10 minutes. Both fecal and blood samples were kept at 4°C during transport to the laboratory and then aliquoted and stored at -80°C until further analysis, as described previously [178]. Blood was tested using the IDEXX MAP Antibody test kit (IDEXX Laboratories, Inc., Westbrook, ME, Product No. 99-14444). The presence of MAP in feces was confirmed through qPCR using the VetMAX™-Gold MAP

Detection Kit (Life Technologies, Corp., Austin, TX, Thermo-Fisher Catalogue No. A29809) following DNA extraction using the ZR-96 Fecal DNA Kit (Zymo Research Corp., Irvine, CA, Zymo Research Catalogue No. D6011).

555 cows from the 3,452 tested in triplicate displayed positive results for MAP infection using both the qPCR and ELISA assays (data not shown). Cows from each herd were selected based on the highest levels, i.e., those samples which had the lowest cycle threshold (Ct) value during qPCR. The mean age of the 67 cows for work described in Chapter IV, at the time of selection for MAP culture was, 5.26 ± 1.60 years. The mean age of the 14 high-shedding cows for work described in Chapter V, at the time of selection for MAP culture, was 5.94 years. The number of MAP bacteria excreted in feces was evaluated by qPCR using standard curves of the MAP K-10 strain (ATCC BAA-968D-5) made with nine serial dilutions (five and two-fold, alternatively) from 10 pg to 0.001 pg (2,000-0.2 genomic copies, respectively), based on a genome size of 4.83 Mb. Standard curves were used to extrapolate the genome copy number in one gram of feces **(Appendix Figure A1.1).**

2.2.2 Fecal Decontamination Procedure

The double incubation method, using hexadecylpyridinium chloride (HPC) and a mixture of antibiotics, was employed to decontaminate fecal samples as described previously **(Figure 2.2)** [349,350]. Briefly, 2.0 g of feces was added to 35 mL (for high shedders) or 15 mL (for moderate and low shedders) of sterile water. Samples were vortexed and were left to sit at room temperature for 30 minutes to allow particles to settle. The supernatant (5 mL) was transferred to 25 mL of 1/2X Brain Heart Infusion

(BHI) broth (Hardy Diagnostics, Santa Maria, CA) containing 0.9% HPC (Sigma-Aldrich, Saint-Louis, MO) and incubated at 37°C for 18-24 hours. After incubation, the sample tubes were centrifuged at 900 x g at room temperature for 30 minutes and, immediately after centrifugation, the supernatant was decanted and the pellet was resuspended in1 mL of 1/2X BHI broth containing 100 µg/mL of vancomycin (US Pharmacopeia, Rockville, MD), 100 µg/mL of nalidixic acid and 50 µg/mL of amphotericin B (Sigma-Aldrich). The sample was incubated at 37°C for 48 hours after which it was transferred to a cryogenic holding tube. The sample was then used to inoculate solid and liquid cultures, with an aliquot stored at -80°C.

2.2.3 Growth of MAP Cultures

To obtain single colonies, slants of Herrold's egg yolk agar with amphotericin, nalidixic acid, vancomycin and mycobactin J (Becton, Dickinson and Company, Franklin Lakes, NJ) were inoculated with up to 100 µL of a decontaminated MAP suspension. Slants were incubated at 37°C and observed every 2-4 weeks up to 25 weeks. Small white colonies usually appeared after 4-6 weeks of incubation, though incubation was continued beyond this point to ensure that single colonies would be collected, which are more visible when they reach > 0.5 mm in diameter. For cattle with high loads of MAP, a 1:10 dilution of the decontamination product was performed to ensure collection of the isolated strain of MAP, as the high-density undiluted product would result in too high of a density to allow for the collection of individual colonies **(Figure 2.3)**.

For low MAP-load samples which failed to grow sufficient colonies on solid media, some of which were analyzed in Chapter IV, an intermediate step in liquid media

between decontamination and growth on an agar slant was required **(Figure 2.3)**. As previously described [351], two 50 mL tubes containing 6 mL of the M7H9C medium were inoculated with 100 and 400 µL of the decontaminated fecal suspension and were incubated at 37°C for up to 25 weeks. As a pellet of egg yolk, which contains MAP, had formed at the bottom of the tube as growth continued, the culture was lightly vortexed weekly. From the 9th week of incubation, then every 4-6 weeks after this point, the growth of MAP was monitored by the Morse staining method. Briefly, 2 µL of the pellet of egg yolk was spread on a microscope slide and was stained using the TB Fluorescent Stain Kit M (Becton, Dickinson, and Company, Franklin Lakes, NJ, SKU: 212519) according to the manufacturer's protocol. Fluorescence of acid-fast bacilli was observed using the Evos FL Auto Microscope (Thermo Fisher Scientific, Waltham, MA) with the GFP fluorescence filter. When the density of MAP (assessed visually) was sufficient, solid cultures were processed as described above. The egg yolk pellet from the M7H9C culture was passed 4-5 times through a 26-gauge needle to dissociate as many MAP clumps as possible. Serial dilutions (1:10, 1:100. 1:1,000, 1:10,000 depending on the observed MAP density) were inoculated into Middlebrook 7H9 medium (Becton, Dickinson, and Company, Franklin Lakes, NJ,) supplemented with 10% Oleic-Albumin-Dextrose-Catalase enrichment (OADC, Becton, Dickinson and Company, Franklin Lakes, NJ,) and 0.05% Tween 80 (Sigma-Aldrich). 10 µL were used to inoculate Herrold's egg yolk agar with amphotericin, nalidixic acid, vancomycin and mycobactin J.

After 9-15 weeks of incubation, selected colonies from each cow were cultured in 1.5 mL of Middlebrook 7H9 broth supplemented with 10% OADC enrichment, 2%

glycerol (Wisent Inc., Saint-Jean-Baptiste, QC, Canada) and 2 mg/L of mycobactin J (Allied Monitor Inc., Fayette, MO) at 37°C with agitation **(Figure 2.3)**. For Chapter IV, a single colony from each of the 67 cattle was selected for this portion of the study, and an axenic culture was made from this single colony to obtain a sufficient amount of MAP (~10^9) for DNA extraction. For Chapter V, instead of selecting a single colony from the 14 animals examined, up to ten colonies were selected, each grown within their own axenic culture to the required concentration. From the 7th week, then every 2-4 weeks, absorbance at 600 nm was measured until it reached 0.7 (corresponding with the mid-log phase of bacterial growth). A portion of the culture was stored at -80°C in a preserving medium containing 5% Tryptic Soy Broth (Sigma-Aldrich) and 30% glycerol, after 15-minute centrifugation at 4,000 x g. The remaining portion was centrifuged for 15 minutes at 14,000 x g and the pellet was stored at -80°C until DNA extraction could be carried out.

2.2.4 DNA Extraction and Sequencing Preparation

Genomic DNA was extracted using the Quick-DNA Fecal/Soil Microbe Miniprep Kit (Zymo Research Corp., Irvine, CA, Zymo Research Catalogue No. D6010) according to the manufacturer's instructions. Homogenization of the MAP pellet was performed using an Omni Bead Ruptor 24 (Omni International Inc., Kennesaw, GA), twice for 1 minute at 6 m/s. DNA concentration was quantified using a NanoDrop One and Qubit 4 Fluorometer (Thermo Fisher Scientific). For the data shown in both Chapters IV and V of this thesis, shotgun libraries from a total of 67 and 139 DNA isolates, respectively, were prepared by the Centre d'expertise et de services Génome Québec (Montreal, QC,

Canada) and sequenced using 150-base pair (bp) paired-end reads with Illumina NovaSeq 6,000 technology (SP flowcell).

2.3 Processing of Sequencing Data for Chapter IV

2.3.1 Initial Read Processing

Several bioinformatic tools were used to process the raw sequencing reads for each of the 67 isolates examined in Chapter IV **(Figure 2.4)**. Fastp v. 0.23.2 was used to trim adapter sequences from the raw FASTQ files and to exclude any unpaired sequence reads that may be present [352]. The trimmed FASTQ files were checked with FastQC to validate read quality [353]. Except for three isolates with suspected contaminants, all results were considered acceptable. Kraken2 v. 2.1.2 was then used to perform taxonomic identification of the reads using a custom-built database [354] using the National Center for Biotechnology Information (NCBI) RefSeq database and 138 MAP sequences from the European Nucleotide Archive (ENA) [355]. KrakenTools v. 1.2 was used to extract reads corresponding to the family Mycobacteriaceae (NCBI TXID 1762) [356], removing any suspected contaminant reads from all samples. The genome assembler SPAdes v. 3.15.4 was run using six kmer sets (21, 33, 55, 77, 99, and 127) with the "isolate" setting to assemble the reads from each isolate into contigs [357,358]. RagTag v. 2.1.0 was used to improve the contig assemblies produced by SPAdes [359]. QUAST v. 5.0.2 was used on both SPAdes contig/scaffold and RagTag scaffold results to identify any outstanding quality control issues [360]. The CheckM v. 1.2.0 lineage-based workflow was used on RagTag scaffold results to examine the completeness and level of contamination present within each of the assemblies [361].

2.3.2 Annotation of MAP genomes

In addition to the 67 field isolates gathered from farms, 10 NCBI reference sequences were also downloaded for comparison. These publicly available genome sequences included that of MAP K-10 (GenBank accession No. NC_002944.2), Telford (No. NZ_CP033688.1), S397 (No. NZ_CP053749.1), MAPK_JB16/15 (No. NZ_CP033911.1), NL 89C (No. NZ_LGRY01000001-NZ_LGRY01000098), NL 93B (No. NZ_LGRZ 01000001-NZ_LGRZ01000090), NL 95A (No. NZ_LGSA01000001-NZ_LGSA01000094), NL 95B (No. NZ_LGSB01000001-NZ_LGSB01 000090), NL 95E (No. NZ_LGSC01000001-NZ_LGSC01000097), and NL 96E (No. NZ_LGSD01000001-NZ_LGSD01000090).

To annotate the assemblies **(Figure 2.4)**, Prokka v. 1.14.5 was run using four of the NCBI sequences to act as trusted annotation files for each of the four different MAP types (representing type I, type II, type III, and type B MAP strains) [362]. Prokka HMM databases were enhanced with Pfam and TIGRFam databases [363,364] to allow additional accurate protein annotations to be added. Prokka was also run for each of the downloaded RefSeq strains using their respective .gff files as a reference.

2.3.3 Variant Analysis

Snippy v. 4.6.0 was used to call variants (including SNPs, insertions, deletions, multi-nucleotide polymorphisms, and complex variations) within both the reference genomes and the FASTQ files processed by trimming and quality assessment for each isolate **(Figure 2.4)** [365]. Snippy was used four times, with each run using a different reference genome corresponding to type I, II, III, and B MAP strain types. The files

produced by the initial Snippy analysis were examined to determine the number, type, and location of variants with respect to each reference strain. For the construction of core SNP phylogenies based on each type of strain, further processing using snippy_core (for identifying and collecting core SNPs), snippy_clean (to remove SNPs of poor quality and any remaining gaps), gubbins v. 3.2.1 (to find and remove any regions which may be indicative of recombination) and SNP-sites v. 2.5.1 (to create the final multiple sequence alignment) was performed [366,367]. Variant patterns and identification of unique SNPs were compared using Geneious Prime v. 2022.1.1 (Biomatters, Inc., USA). Statistical analysis was performed by analysis of variance (ANOVA) with unequal variances. A comparison of the herds was done with a Tukey correction. For the association with herd prevalence, Spearman Correlation analysis was performed in Statistical Analysis System (SAS, Release 9.4, 2002-2012. SAS Institute Inc., Cary, NC).

2.3.4 Phylogeny Construction

Data on the variant analysis allowed for the construction of four core SNP phylogenies. Each core SNP tree was created with IQ-TREE v. 1.6.12 which selected the optimal model for tree building (GTR+ASC) and built each tree with 1,000 bootstraps [368,369]. Tree visualization was performed using ITOL v. 6.0 [370].

2.3.5 MIRU-VNTR and ML-SSR Typing

Geneious Prime v. 2022.1.1 (Biomatters, Inc., USA) was used to extract *in-silico* ML-SSR results from the assemblies [286]. If the repeat number at an SSR locus was unclear, the individual reads were examined to validate the number of repeats. MIRU-VNTR repeats were counted using the Tandem Repeats Finder [287,371]. MIRU-VNTR and

ML-SSR data were reported in a datasheet file alongside the isolate number and herd ID **(Table 2.2)**. Importing the data files into the Phandango visualization software, alongside the type II phylogeny and shedding data, allowed for a visual representation of the *in-silico* analysis [372]. ML-SSR and MIRU-VNTR patterns were compared with those found in the MAC-INMV database to compare analyzed strains with previously characterized isolates [373]. The Hunter-Gaston discriminatory index (DI) [374], was used to calculate the discriminatory power of both the ML-SSR and MIRU-VNTR loci. The DI was calculated using the following equation:

$$DI = 1 - \left[\frac{1}{N(N-1)}\sum_{j=1}^{s} n_j(n_j - 1)\right]$$

where N is the total number of isolates being typed, s is the total number of distinct types being discriminated by the respective typing method, and n_j is the number of isolates belonging to the jth type.

The MIRU-VNTR results for the 67 field isolates were confirmed with PCR using 5 µL Green GoTaq buffer (Promega), 1.5 µL 25 mM $MgCl_2$ (Promega), 2.5 µL 2 mM dNTPs (Promega), 0.1 µL (5 U/µL) GoTaq G2 Flexi DNA polymerase (Promega), 2.5 µL (10 µM) forward and reverse primers (Eurogentec), and 5 µL DNA (≥ 10 ng/µL). A volume of 1 µL of dimethylsulfoxide (DMSO) (Sigma D2650) and/or 5 µL of 5 M betaine (Sigma B0300) were added as indicated [287] and the total reaction volume was completed to 25 µL with molecular grade distilled water. The PCR was carried out using a Techne TC512 thermocycler according to the following reaction protocol: an initial incubation of 5 minutes at 94°C, followed by 40 cycles of 30 seconds at 94°C, 30 seconds at the

temperature of specific hybridization of each locus, 30 seconds at 72°C, and a final cycle of 7 minutes at 72°C. PCR reaction products were analyzed by electrophoresis on a 1.5% agarose gel for 1 hour at 100 V to determine the number of repeats present at each of the eight loci.

2.4 Processing of Sequencing Data for Chapter V

2.4.1 Genome Assembly and Annotation, Phylogeny Construction and Strain Typing

Analysis of the 139 isolates described in Chapter V was conducted similarly to the procedures performed for those in Chapter IV with a few key differences designed to streamline the procedure **(Figure 2.5)**. Both genome annotation and SNP analysis were performed using only the K-10 reference sequence (NC_002944.2), and snippy_core, snippy_clean, and gubbins were not used for variant analysis. Instead, the "consensus.subs.fa" FASTA files generated for each isolate by Snippy were combined into a single FASTA file, which was directly processed using SNP-sites. The ".vcf" files produced by Snippy for each isolate were mapped onto the K-10 genome sequence using Geneious Prime v. 2022.1.1 (Biomatters, Inc., USA), allowing for the visualization of variants such as indels and SNPs in the context of the reference genome. Variant analysis also allowed for the construction of SNP-based phylogenies containing all isolates from the 14 animals, as well as phylogenies for isolates from each animal. Each core SNP tree was created with IQ-TREE v. 1.6.12 which selected the optimal model for tree building in each case **(Table 2.3)** Analysis of the 8 MIRU-VNTR loci and 11 ML-SSR loci was performed using the same *in-silico* protocol described for Chapter IV **(Table 2.4)**, however, no PCR of MIRU-VNTR loci was performed for Chapter V.

2.4.2 Prevalence of variable reading frames in SSR Loci 1 and 2

In addition to the 139 MAP genomes examined as part of Chapter V of this thesis, the sequences of 53 other isolates from our collection (those from Chapter IV) or publicly available in the NCBI database (n=41) **(Appendix Table A1)** were used for analyzing the prevalence of different nucleotide repeats associated with SSR loci 1 and 2 (henceforth referred to as simply SSR1 and SSR2, respectively) [286]. The analysis included 1,388 sequences from the NCBI Sequence Read Archive (SRA) as of December 14, 2022). FASTQ files corresponding to each record were downloaded and compressed using the "prefetch" and "fasterq-dump" commands from the SRA-Toolkit v. 3.0.2 [375] along with the "gzip" Unix command. Cleaning and assembly of reads from the SRA and other NCBI records were performed using Fastp v. 0.23.2 and SPAdes v. 3.15.4, respectively [352,357,358]. Sequence selection and extraction of the information at the loci SSR1 and SSR2 were performed using the *in-silico*-PCR software package [376]. These shortened sequences were examined using Geneious Prime v. 2022.1.1 (Biomatters, Inc., USA), allowing for the compilation of SSR1 and SSR2 repeat numbers and the determination of the reading frame of the gene through examination of the number of variable repeats **(Appendix Table A1)**. In some instances, the sequence from the SRA records had to be assembled before analysis, to obtain unambiguous results.

2.4.3 Predictive modelling of proteins encoded by genes containing SSR1 and SSR2

Loci listed as proteins WP_010949291.1 and WP_134797017.1 in the K-10 GenBank accession (NC_002944.2), referred to as ORF1 and ORF2 in this publication, contain the SSR1 and SSR2 repeats, respectively. Statistical analysis of both the reading

frames and SSR repeat sizes in *ORF1* and *ORF2* was performed using a χ^2 test of independence using a p-value threshold of 0.05. The most prevalent repeat sizes for each of the three reading frames within these two loci were identified, and the complete nucleotide sequences of the loci containing these variants were extracted and translated into their amino acid sequences using Geneious Prime v. 2022.1.1 (Biomatters, Inc., USA). BLASTn searches filtered to detect DNA sequences with ≥80% percent identity and query coverage were used for detecting orthologous genes [377]. Representative amino acid sequences of ORF1 and ORF2 from MAP in each reading frame were also queried using BLASTp [377]. Conserved domains (if present) were detected using the NCBI conserved domain tool [378]. SignalP v. 6.0 was used to detect the presence of signal peptides in the predicted proteins [379], whereas ColabFold v. 1.5.2 [380] was used for protein structure prediction using Alphafold2 along with MMseqs2 with default settings [381,382]. Protein structures were visualized in ChimeraX [383,384].

2.5 Figures and Tables

2.5.1 Figures

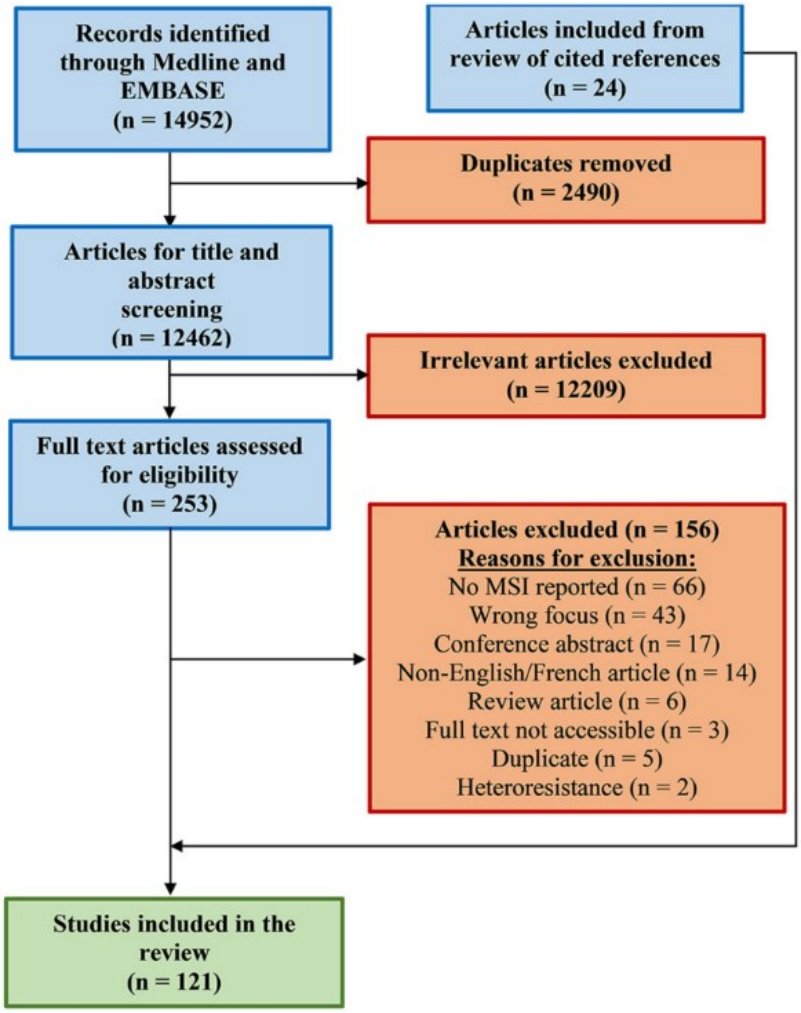


Figure 2.1: Flow diagram of the search strategy. Modified preferred reporting items for systematic reviews and meta-analysis (PRISMA) results for a search on mycobacterial MSI detection. Additional details are provided in Section 2.1 and **Table 2.1**

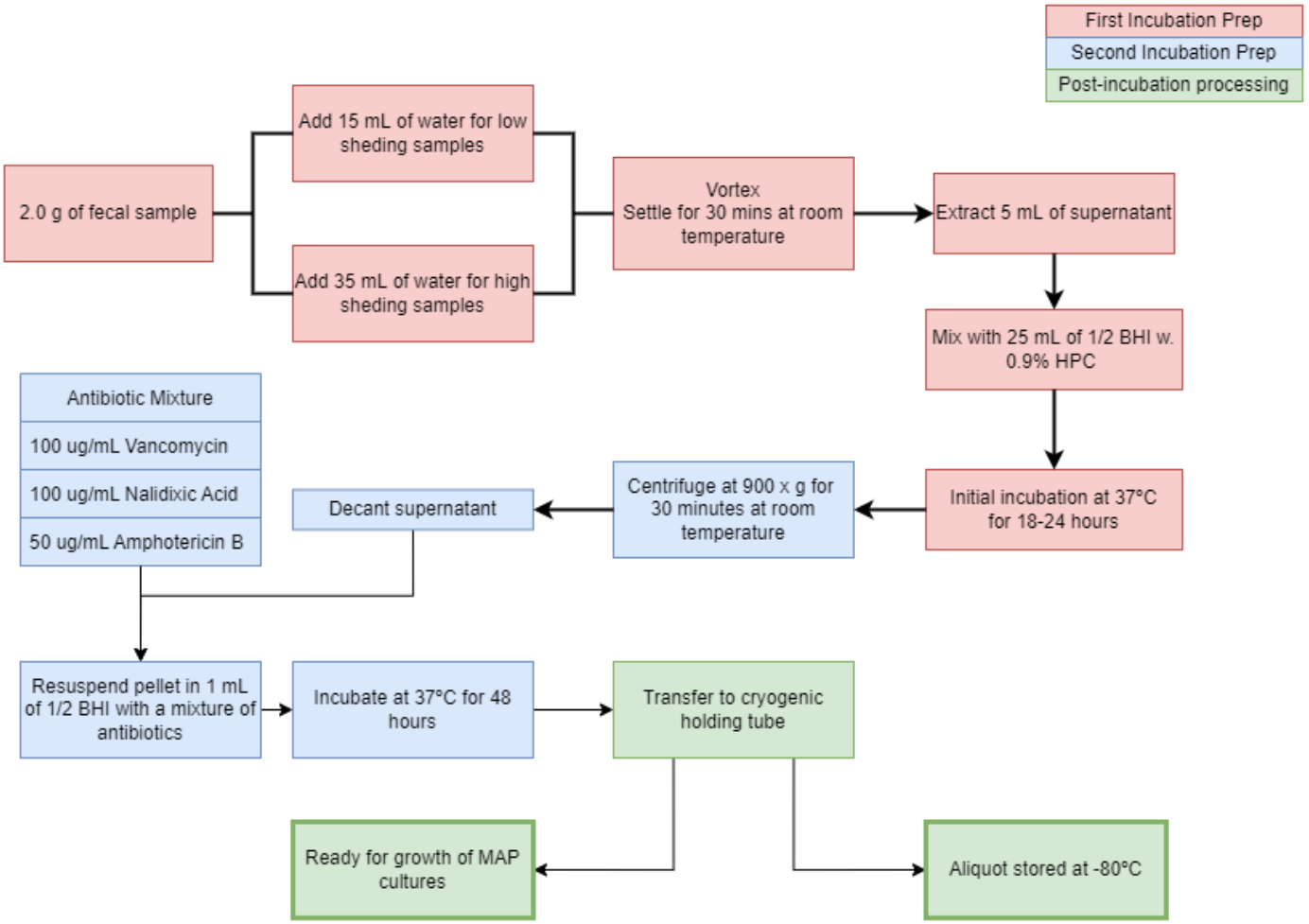


Figure 2.2: Procedure followed to decontaminate fecal samples and prepare samples for MAP culturing. Further details are provided in Section 2.2.

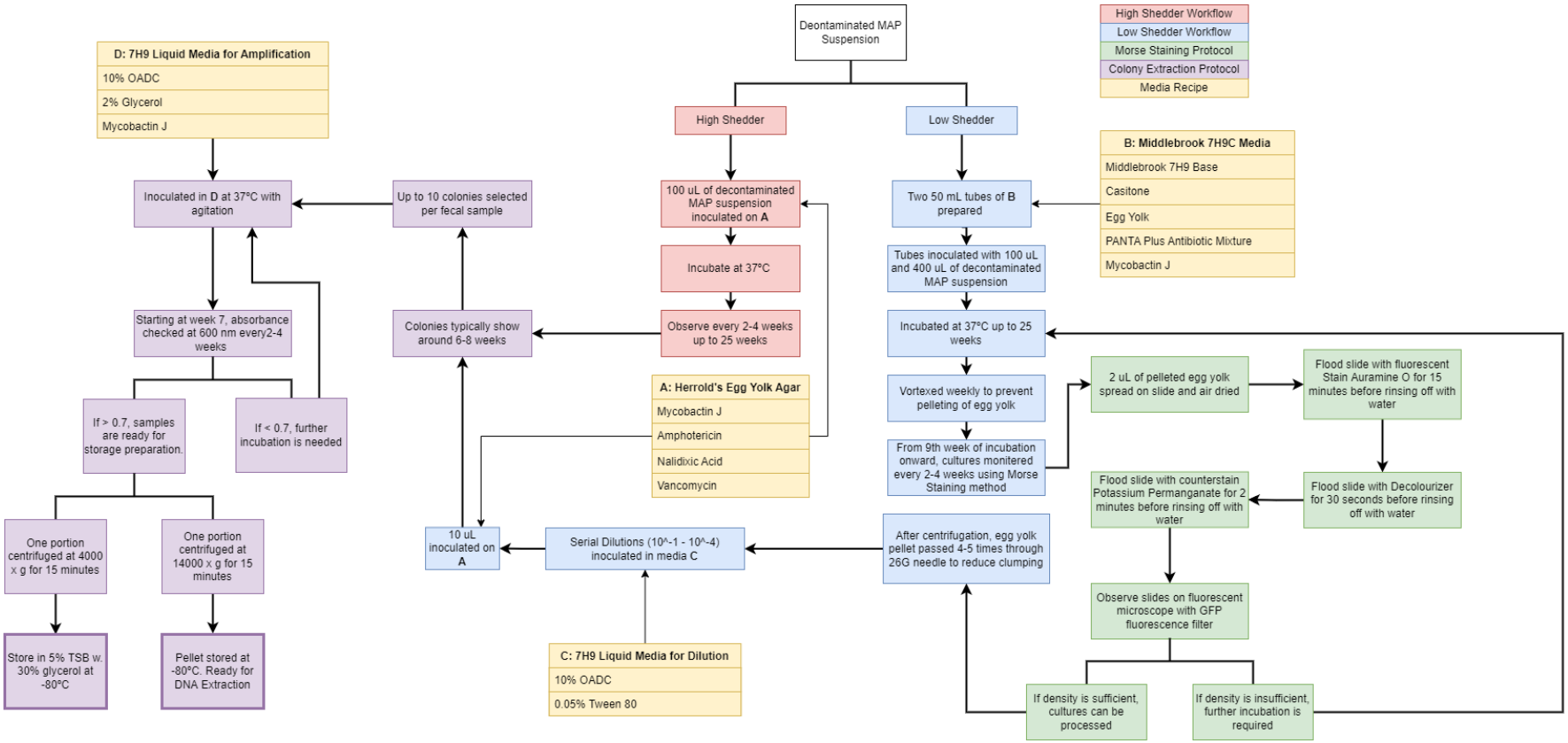


Figure 2.3: Pipeline of the procedure used to culture MAP isolates from both high and low-shedding cattle. Further details are provided in Section 2.2.

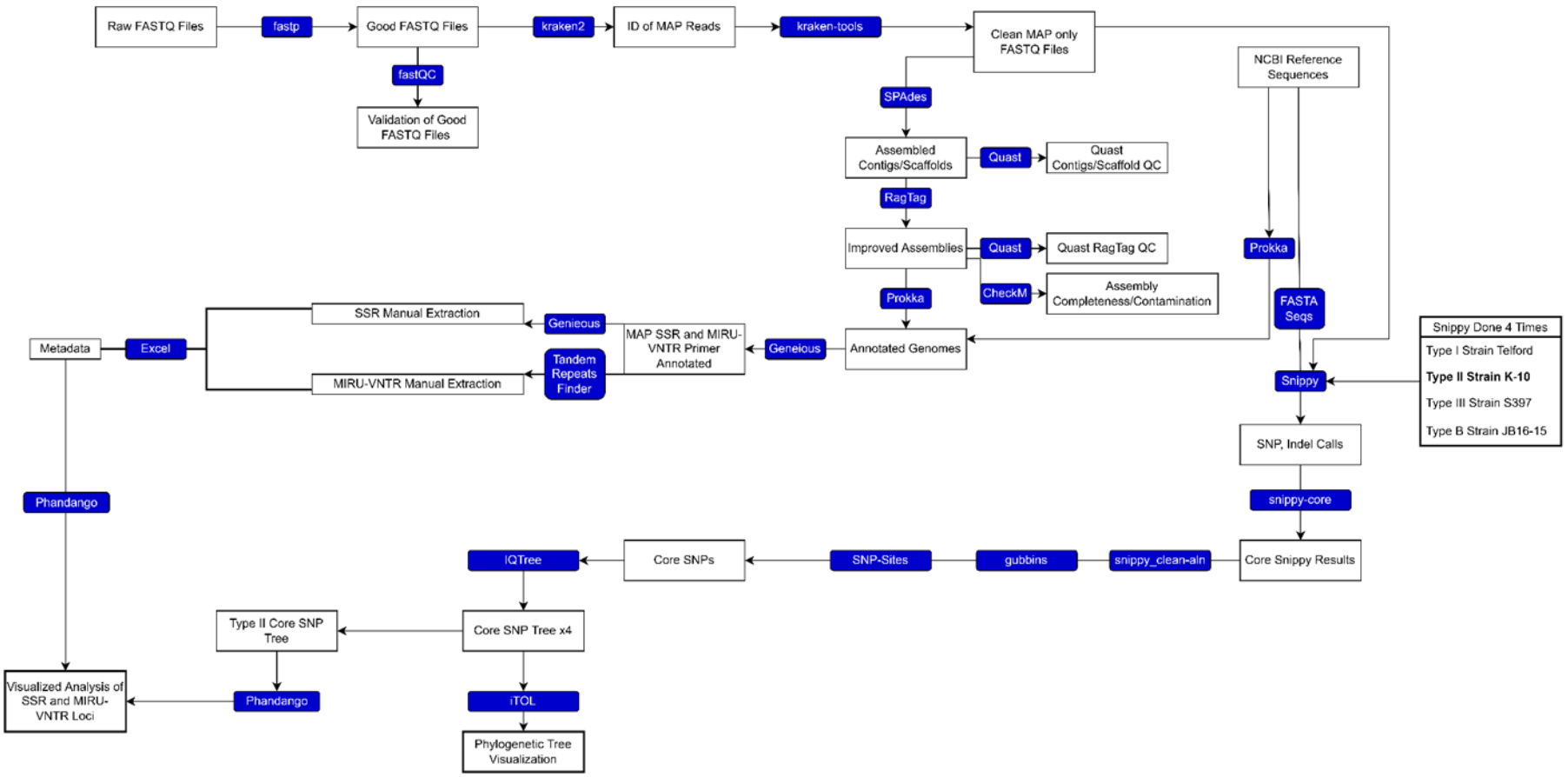


Figure 2.4: **Pipeline of bioinformatic tools (blue) and files (white) used to analyze the whole genome sequencing data in Chapter IV of this thesis.** The exact parameters used for each tool are described in Section 2.3.

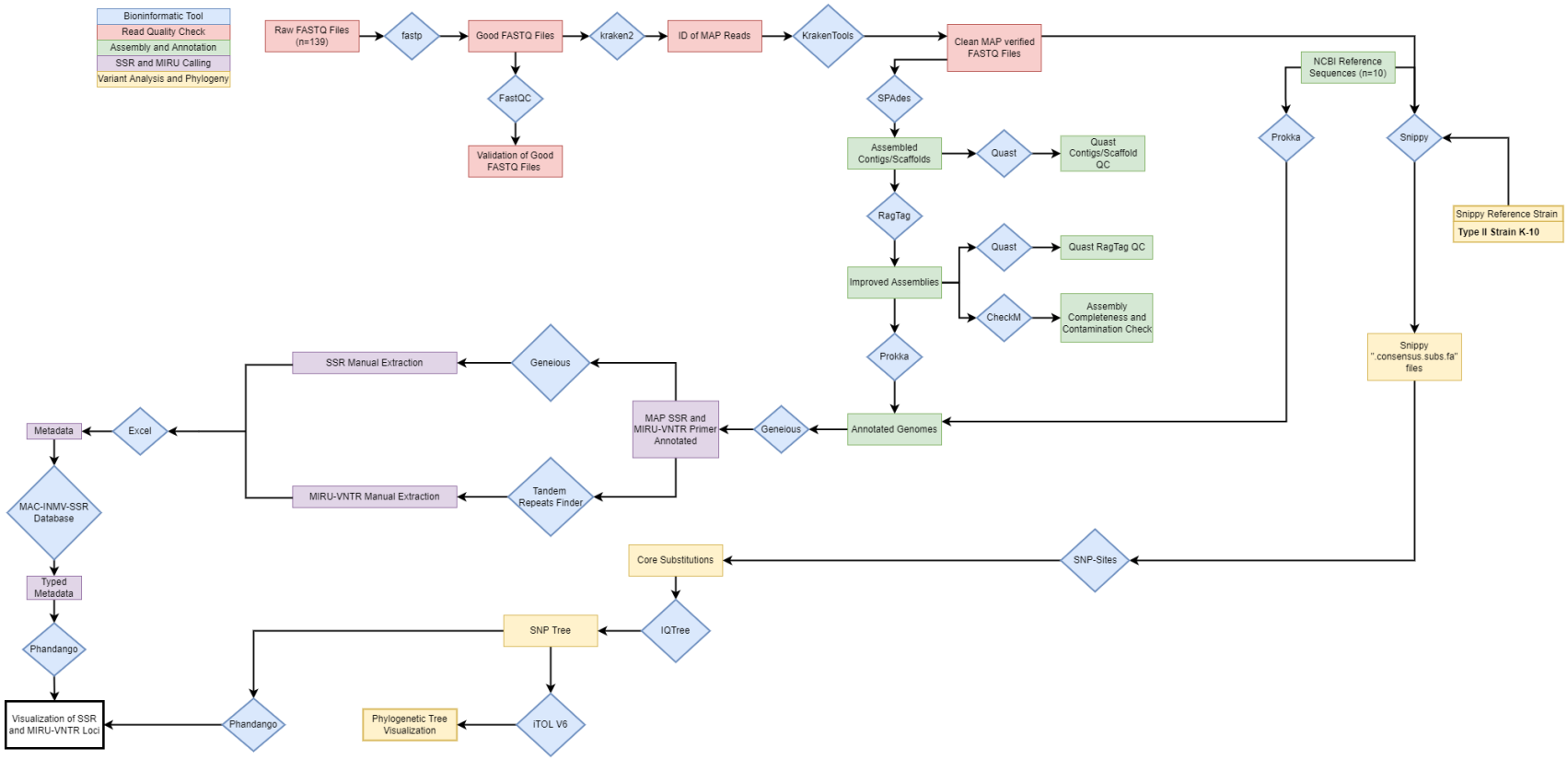


Figure 2.5: Pipeline of bioinformatic tools (diamonds) and files (rectangles) used to analyze the whole genome sequencing data in Chapter V of this thesis. The exact parameters used for each tool are described in Section 2.4.

2.5.2 Tables

Table 2.1: Keywords, index terms, and strings used to search databases for reports on mycobacterial MSIs. Details regarding term selection are described in Section 2.1.

A: Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) 1946 to June 9, 2020: Searched June 11, 2020.

#	Search string	Results
1	exp Whole Genome Sequencing/	7158
2	exp Bacteriophage Typing/	4116
3	((Bacteriophage or phage) adj3 typing).ab,ti,kf.	2182
4	IS6110 RFLP.ab,ti,kf.	261
5	MIRU-VNTR.ab,ti,kf.	649
6	Mycobacterial interspersed repetitive unit-variable number tandem repeat typing.ab,ti,kf.	31
7	VNTR.ab,ti,kf.	4993
8	varia* number* tandem* rep* typ*.ab,ti,kf.	91
9	MIRU.ab,ti,kf.	817
10	Mycobacter* interspers* rep* unit*.ab,ti,kf.	614
11	(molecular adj2 (detect* or typing or method* or technique* or diagnos*)).ab,ti,kf.	72323
12	whole genom* sequenc*.ab,ti,kf.	18071
13	(assay adj2 (genom* or molecul*)).ab,ti,kf.	3169
14	typing.ab,ti.	51409
15	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14	149327
16	exp Mycobacterium/	91603
17	exp Mycobacterium Infections, Nontuberculous/	34845
18	exp Tuberculosis/	191333
19	Coinfection/	10498
20	"Mycobacteri* tuberculosis".ab,ti.	48894
21	Mycobacter*.ab,ti,kf.	97999
22	"m. tuberculosis".ab,ti.	18616
23	"m. africanum".ab,ti.	267
24	"m. bovis".ab,ti.	4581
25	"mycobacteri* bovis".ab,ti.	7148
26	"m. caprae".ab,ti.	98
27	"m. orygis".ab,ti.	11
28	"m. simiae".ab,ti.	211
29	"M. abscessus".ab,ti.	1122
30	"M. chelonae".ab,ti.	616
31	"M. fortuitum".ab,ti.	1132
32	"mycobacteri* gordonae".ab,ti.	320
33	"M. intracellulare".ab,ti.	985
34	"m. avium".ab,ti.	3710
35	"mycobacteri* avium".ab,ti.	9189

36	"m. leprae".ab,ti.	3013
37	"mycobacteri* leprae".ab,ti.	4383
38	Nontuberculous.ab,ti.	4118
39	"heteroresistan*".ab,ti.	580
40	coinfection.ab,ti.	9484
41	co-infection.ab,ti.	10853
42	((mixed or multiple or double or simultaneous or different) adj2 (infection* or strain*)).ab,ti.	50700
43	16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41 OR 42	356175
44	15 AND 43	8932

B: Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Daily and Versions(R) 1946 to August 24, 2020: Revised search to polyclonal and only retrieve results including this new term: Searched August 25, 2020

#	Search string	Results
1	exp Whole Genome Sequencing/	8018
2	exp Bacteriophage Typing/	4117
3	((Bacteriophage or phage) adj3 typing).ab,ti,kf.	2187
4	IS6110 RFLP.ab,ti,kf.	260
5	MIRU-VNTR.ab,ti,kf.	654
6	Mycobacterial interspersed repetitive unit-variable number tandem repeat typing.ab,ti,kf.	31
7	VNTR.ab,ti,kf.	5024
8	varia* number* tandem* rep* typ*.ab,ti,kf.	91
9	MIRU.ab,ti,kf.	822
10	Mycobacter* interspers* rep* unit*.ab,ti,kf.	616
11	(molecular adj2 (detect* or typing or method* or technique* or diagnos*)).ab,ti,kf.	73468
12	whole genom* sequenc*.ab,ti,kf.	18763
13	(assay adj2 (genom* or molecul*)).ab,ti,kf.	3221
14	typing.ab,ti.	50406
15	polyclonal:ti,ab,kw	52613
16	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15	200497
17	exp Mycobacterium/	92185
18	exp Mycobacterium Infections, Nontuberculous/	34993
19	exp Tuberculosis/	192193
20	Coinfection/	10824
21	"Mycobacteri* tuberculosis".ab,ti.	45244
22	Mycobacter*.ab,ti,kf.	98848
23	"m. tuberculosis".ab,ti.	18600
24	"m. africanum".ab,ti.	267
25	"m. bovis".ab,ti.	4617
26	"mycobacteri* bovis".ab,ti.	6847
27	"m. caprae".ab,ti.	101
28	"m. orygis".ab,ti.	12
29	"m. simiae".ab,ti.	213
30	"M. abscessus".ab,ti.	1131
31	"M. chelonae".ab,ti.	619
32	"M. fortuitum".ab,ti.	1138
33	"mycobacteri* gordonae".ab,ti.	322
34	"M. intracellulare".ab,ti.	984
35	"m. avium".ab,ti.	3710
36	"mycobacteri* avium".ab,ti.	9163
37	"m. leprae".ab,ti.	3008
38	"mycobacteri* leprae".ab,ti.	4270
39	Nontuberculous.ab,ti.	3970

40	"heteroresistan*".ab,ti.	574
41	coinfection.ab,ti.	9286
42	co-infection.ab,ti.	10635
43	((mixed or multiple or double or simultaneous or different) adj2 (infection* or strain*)).ab,ti.	50700
44	17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 OR 30 OR 31 OR 32 OR 33 OR 34 OR 35 OR 36 OR 37 OR 38 OR 39 OR 40 OR 41 OR 42 OR 43	358223
45	16 AND 44	9825
46	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14	148522
47	44 AND 46	8930
48	45 NOT 47	895

C: EMBASE (Elsevier): Searched June 11, 2020		
#	Search string	Results
1	'mycobacterium'/exp	125,262
2	'mycobacteriosis'/exp	310,109
3	'tuberculosis'/exp	264,147
4	'mixed infection'/exp	35,894
5	'mycobacteri* tuberculosis':ti,ab,kw	54,301
6	mycobacter*:ti,ab,kw	113,620
7	'm. tuberculosis':ti,ab,kw	22,305
8	'm. bovis':ti,ab,kw	4,882
9	'm. caprae':ti,ab,kw	89
10	'm. leprae':ti,ab,kw	3,742
11	nontuberculous:ti,ab,kw	6,133
12	heteroresistan*:ti,ab,kw	703
13	'mixed infection':ti,ab,kw	4,176
14	'm. canetti':ti,ab,kw	13
15	'm. ulcerans':ti,ab,kw	648
16	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15	388,584
17	'whole genome sequencing'/exp	17,608
18	'bacteriophage typing'/exp	3,993
19	'mycobacterial interspersed repetitive unit-variable number tandem repeat typing':ti,ab,kw	28
20	'miru vntr':ti,ab,kw	771
21	vntr:ti,ab,kw	6,275
22	'varia* number* tandem* rep* typ*':ti,ab,kw	92
23	'mycobacter* interspers* rep* unit*':ti,ab,kw	656
24	'whole genom* sequenc*':ti,ab,kw	22,508
25	(molecular NEAR/2 (detect* OR typing OR method* OR technique* OR diagnos*)):ti,ab,kw	96,957
26	(assay NEAR/2 (genom* OR molecul*)):ti,ab,kw	4,779
27	((bacteriophage OR phage) NEAR/3 typing):ti,ab,kw	2,159
28	typing:ti,ab,kw	68,129
29	miru:ti,ab,kw	983
30	#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29	194,819
31	#16 OR #30	7,501

D: EMBASE (Elsevier): Revised search to polyclonal and only retrieve results including this new term: Searched August 25, 2020

#	Search string	Results
1	'mycobacterium'/exp	126,409
2	'mycobacteriosis'/exp	312,545
3	'tuberculosis'/exp	266,260
4	'mixed infection'/exp	36,795
5	'mycobacteri* tuberculosis':ti,ab,kw	54,980
6	mycobacter*:ti,ab,kw	114,906
7	'm. tuberculosis':ti,ab,kw	22,634
8	'm. bovis':ti,ab,kw	4,985
9	'm. caprae':ti,ab,kw	97
10	'm. leprae':ti,ab,kw	3,761
11	nontuberculous:ti,ab,kw	9,244
12	heteroresistan*:ti,ab,kw	762
13	'mixed infection':ti,ab,kw	4,214
14	'm. canetti':ti,ab,kw	13
15	'm. ulcerans':ti,ab,kw	655
16	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15	422,591
17	'whole genome sequencing'/exp	18,909
18	'bacteriophage typing'/exp	3,998
19	'mycobacterial interspersed repetitive unit-variable number tandem repeat typing':ti,ab,kw	30
20	'miru vntr':ti,ab,kw	781
21	vntr:ti,ab,kw	6,340
22	'varia* number* tandem* rep* typ*':ti,ab,kw	94
23	'mycobacter* interspers* rep* unit*':ti,ab,kw	661
24	'whole genom* sequenc*':ti,ab,kw	23,411
25	(molecular NEAR/2 (detect* OR typing OR method* OR technique* OR diagnos*)):ti,ab,kw	98,652
26	(assay NEAR/2 (genom* OR molecul*)):ti,ab,kw	4,866
27	((bacteriophage OR phage) NEAR/3 typing):ti,ab,kw	2,161
28	typing:ti,ab,kw	68,784
29	miru:ti,ab,kw	993
30	polyclonal:ti,ab,kw	62,078
31	#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30	259,744
32	#16 AND #31	9628
33	#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29	195,592
34	#16 AND #33	8917
35	#32 NOT #34	711

Table 2.2: ML-SSR and MIRU-VNTR patterns identified within select isolates extracted from multiple herds in Quebec and Ontario (n=77) as discussed in Chapter IV.

Strain ID[a]	Herd ID[b]	Province	Shedding Status	SSR1[c]	SSR2[c]	SSR3[c]	SSR4[c]	SSR5[c]	SSR6[c]	SSR7[c]	SSR8[c]	SSR9[c]	SSR10[c]	SSR11[c]	INMV SSR Type[d]	292[c]	X3[c]	25[c]	47[c]	3[c]	7[c]	10[c]	32[c]	INMV MIRU Type[d]
117	ON-1	ON	High	15	13	5	5	5	5	6	5	5	5	5	N/P 12a	3	2	3	3	2	2	2	8	INMV 2
137	ON-1	ON	High	20	10	5	5	5	5	6	5	5	5	5	N/P 15	3	2	3	3	2	2	2	8	INMV 2
233	ON-1	ON	High	19	13	5	5	5	5	5	5	5	5	5	MLSSR 93e	3	2	3	3	2	2	2	8	INMV 2
236	ON-1	ON	High	14	13	5	5	5	5	5	5	5	5	5	MLSSR 93b	3	2	3	3	2	2	2	8	INMV 2
242	ON-1	ON	High	22	13	5	5	5	5	6	5	5	5	5	N/P 12b	3	2	3	3	2	2	2	8	INMV 2
252	ON-1	ON	High	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10d	3	2	3	3	2	2	2	8	INMV 2
211	ON-2	ON	Moderate	18	11	5	5	5	5	5	5	5	5	5	MLSSR 8b	3	2	3	3	2	2	1	8	INMV 3
552	ON-2	ON	Moderate	13	9	5	5	5	5	5	5	5	5	5	MLSSR 7a	3	2	3	3	2	2	1	8	INMV 3
560	ON-2	ON	High	11	9	5	5	5	5	4	5	5	5	5	N/P 8	3	2	3	3	2	2	1	8	INMV 3
1452	ON-2	ON	High	7	10	5	5	5	4	5	4	4	5	5	MLSSR 28	3	2	3	3	2	2	2	8	INMV 2
325	ON-3	ON	Moderate	7	11	5	5	5	5	5	5	5	5	5	MLSSR 17	3	2	3	3	2	2	2	8	INMV 2
326	ON-3	ON	High	7	10	5	5	5	5	5	5	5	5	5	MLSSR 11	3	2	3	3	2	2	2	8	INMV 2
1240	ON-3	ON	Low	7	14	5	5	5	5	5	5	5	5	5	N/P 3c	3	2	3	3	2	2	2	8	INMV 2
1402	ON-3	ON	High	7	10	5	5	5	5	5	5	5	5	5	MLSSR 11	3	2	3	3	2	2	2	8	INMV 2
342	ON-4	ON	Moderate	9	UNK	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
345	ON-4	ON	Low	7	12	5	5	5	5	5	5	5	5	5	N/P 3a	3	2	3	3	2	2	2	8	INMV 2
863	ON-4	ON	High	9	17	5	5	5	5	5	5	5	5	5	N/P 6b	3	2	3	3	2	2	2	8	INMV 2
464	ON-5	ON	Moderate	15	12	5	5	5	5	5	5	5	5	5	MLSSR 93c	3	2	3	3	2	2	2	8	INMV 2
897	ON-5	ON	High	21	10	5	5	5	5	5	5	5	5	5	MLSSR 10f	3	2	3	3	2	2	2	8	INMV 2
1409	ON-5	ON	High	11	12	5	5	5	5	5	5	5	5	5	N/P 9a	3	2	3	3	2	2	2	8	INMV 2
586	ON-6	ON	Moderate	14	12	5	5	5	5	5	5	5	5	5	MLSSR 93a	3	2	3	3	2	3	2	8	INMV 8
594	ON-6	ON	High	7	13	5	5	5	5	4	5	5	5	5	N/P 4a	3	2	3	3	2	2	2	8	INMV 2
1007	ON-6	ON	High	7	12	5	5	5	5	5	5	5	5	5	N/P 3a	3	2	3	3	2	2	2	8	INMV 2
1023	ON-6	ON	High	7	14	5	5	5	5	4	5	5	5	5	N/P 4b	3	2	3	3	2	2	2	8	INMV 2
1495	ON-6	ON	High	9	11	5	5	5	5	6	5	5	5	5	N/P 5	3	2	3	3	2	2	2	8	INMV 2
1512	ON-6	ON	High	7	10	5	5	5	5	4	5	5	5	5	N/P 1	3	2	3	3	2	2	2	8	INMV 2
631	ON-7	ON	Low	7	10	5	5	5	5	4	5	5	5	5	N/P 1	3	2	3	3	2	2	2	8	INMV 2
638	ON-7	ON	Moderate	20	10	5	5	5	5	5	5	5	5	5	MLSSR 10e	3	2	3	3	2	2	2	8	INMV 2

648	ON-7	ON	Low	UNK	10	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
910	ON-8	ON	High	7	13	5	5	5	5	4	5	5	5	5	N/P 4a	3	2	3	3	2	2	2	8	INMV 2
918	ON-8	ON	High	7	11	5	5	5	5	4	5	5	5	5	N/P 2	3	2	3	3	2	2	2	8	INMV 2
928	ON-8	ON	High	11	9	5	5	5	5	5	5	5	5	5	N/P 16	3	2	3	3	2	2	2	8	INMV 2
1419	ON-8	ON	High	7	11	5	5	5	5	4	5	5	5	5	N/P 2	3	2	3	3	2	2	2	8	INMV 2
23	QC-1	QC	High	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10a	3	2	3	3	2	2	2	8	INMV 2
32	QC-1	QC	High	11	10	5	5	5	5	5	5	5	5	5	MLSSR 38	3	2	3	3	2	2	2	8	INMV 2
219	QC-1	QC	High	11	10	5	5	5	5	5	5	5	5	5	MLSSR 38	3	2	3	3	2	2	2	8	INMV 2
366	QC-1	QC	Moderate	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10a	3	2	3	3	2	2	2	8	INMV 2
160	QC-2	QC	High	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10d	3	2	3	3	2	2	2	8	INMV 2
183	QC-2	QC	High	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8a	3	2	3	3	2	2	2	8	INMV 2
191	QC-2	QC	High	14	10	5	5	5	5	5	5	5	5	5	MLSSR 10b	3	2	3	3	2	2	2	8	INMV 2
171	QC-3	QC	High	17	10	5	5	5	5	5	5	5	5	5	MLSSR 10c	3	2	3	3	2	2	2	8	INMV 2
206	QC-3	QC	High	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8a	3	2	3	3	2	2	2	8	INMV 2
418	QC-3	QC	Moderate	20	10	5	5	5	5	5	5	5	5	5	MLSSR 10e	3	2	3	3	2	2	2	8	INMV 2
375	QC-4	QC	High	18	10	5	5	5	5	4	5	5	5	5	N/P 14	3	2	3	3	2	2	1	8	INMV 3
400	QC-4	QC	High	20	11	5	5	5	5	4	5	5	5	5	N/P 10b	3	2	3	3	2	2	1	8	INMV 3
869	QC-4	QC	High	13	11	5	5	5	5	4	5	5	5	5	N/P 10a	3	2	3	3	2	2	1	8	INMV 3
878	QC-4	QC	High	15	16	5	5	5	5	4	5	5	5	5	N/P 13	3	2	3	3	2	2	1	8	INMV 3
392	QC-5	QC	Moderate	7	9	5	5	5	5	5	5	5	5	5	MLSSR 13	3	2	3	3	2	2	2	8	INMV 2
885	QC-5	QC	High	7	9	5	5	5	5	5	5	5	5	5	MLSSR 13	3	2	3	3	2	2	2	8	INMV 2
1071	QC-5	QC	Moderate	7	9	5	5	5	5	5	5	5	5	5	MLSSR 13	3	2	3	3	2	2	2	8	INMV 2
441	QC-6	QC	Low	14	9	5	5	5	5	4	5	5	5	5	N/P 11	3	2	3	3	2	2	1	8	INMV 3
500	QC-7	QC	Moderate	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10d	3	2	3	3	2	2	1	8	INMV 3
505	QC-7	QC	Moderate	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10d	3	2	3	3	2	2	1	8	INMV 3
1144	QC-7	QC	Moderate	18	9	5	5	5	5	5	5	5	5	5	MLSSR 7b	3	2	3	3	2	2	1	8	INMV 3
508	QC-8	QC	Moderate	7	10	5	5	5	5	5	5	5	5	5	MLSSR 11	3	2	3	3	2	2	2	8	INMV 2
1157	QC-8	QC	Moderate	7	13	5	5	5	5	5	5	5	5	5	N/P 3b	3	2	3	3	2	2	2	8	INMV 2
1438	QC-8	QC	High	7	10	5	5	5	5	5	5	5	5	5	MLSSR 11	3	2	3	3	2	2	2	8	INMV 2
515	QC-9	QC	Moderate	9	11	5	5	5	5	5	5	5	5	5	MLSSR 16	2	2	3	3	2	2	2	8	INMV 13
516	QC-9	QC	Moderate	9	13	5	5	5	5	5	5	5	5	5	N/P 6a	2	2	3	3	2	2	2	8	INMV 13
1427	QC-9	QC	High	10	12	5	5	5	5	5	5	5	5	5	MLSSR 58	2	2	3	3	2	2	2	8	INMV 13

531	QC-10	QC	Moderate	21	10	5	5	5	5	5	5	5	5	5	MLSSR 10f	3	2	3	3	2	2	2	8	INMV 2
938	QC-10	QC	High	18	12	5	5	5	5	5	5	5	5	5	MLSSR 93d	3	2	3	3	2	2	2	8	INMV 2
948	QC-10	QC	High	11	17	5	5	5	5	5	5	5	5	5	N/P 9b	3	2	3	3	2	2	2	8	INMV 2
1032	QC-11	QC	Moderate	7	9	5	5	5	5	5	5	5	5	5	MLSSR 13	3	2	3	3	2	2	2	8	INMV 2
1044	QC-11	QC	High	7	11	5	5	5	5	5	5	5	5	5	MLSSR 17	3	2	3	3	2	2	2	8	INMV 2
1517	QC-11	QC	High	7	9	5	5	5	5	5	5	5	5	5	MLSSR 13	3	2	3	3	2	2	2	8	INMV 2
1578	QC-12	QC	High	10	16	5	5	5	5	4	5	5	5	5	N/P 7	3	2	3	3	2	2	1	8	INMV 3
K-10	NCBI	N/A	N/A	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10d	3	2	3	3	2	2	1	8	INMV 3
MAPK_JB16/15	NCBI	N/A	N/A	7	12	5	5	5	4	4	4	4	5	5	MLSSR 29	4	2	3	3	2	2	1	UNK	UNK
S397	NCBI	N/A	N/A	5	13	4	5	UNK	4	5	3	4	UNK	4	UNK	7	1	3	3	1	1	1	UNK	UNK
Telford	NCBI	N/A	N/A	10	13	4	5	UNK	4	5	3	4	UNK	4	UNK	4	1	3	3	1	1	1	8	INMV 72
NL-89C	NCBI	NL	N/A	7	11	5	5	5	4	5	4	4	5	5	MLSSR 3	2	2	5	3	2	2	2	8	INMV 68
NL-93B	NCBI	NL	N/A	11	11	5	5	5	5	5	5	5	5	5	MLSSR 54	UNK	2	3	3	2	2	2	UNK	UNK
NL-95A	NCBI	NL	N/A	10	11	5	5	5	5	5	5	5	5	5	MLSSR 52	4	2	3	3	2	2	1	UNK	UNK
NL-95B	NCBI	NL	N/A	7	11	5	5	5	5	5	5	5	5	5	MLSSR 17	4	2	3	3	2	2	2	2	N/P 1
NL-95E	NCBI	NL	N/A	7	11	5	5	5	5	5	5	5	5	5	MLSSR 17	4	2	3	3	2	2	1	UNK	UNK
NL-96E	NCBI	NL	N/A	11	10	5	5	5	5	5	5	5	5	5	MLSSR 38	4	2	3	3	2	2	2	UNK	UNK
Discriminatory Index (DI) Value				0.8508	0.8005	0.0000	0.0000	0.0000	0.0299	0.424242	0.0299	0.0299	0.0000	0.0000	0.9846	0.0868	0.0000	0.0000	0.0000	0.0000	0.0299	0.2985	0.0000	0.3740

[a]Typing of isolates from Newfoundland Canada (NL-89C/93B/95A/95B/96E) have been previously published [289,315].

[b]Values from the NCBI database are not included in DI calculations.

[c]Values listed as UNK are not included in DI calculations.

[d]N/P = New Patterns. Patterns are not recognized as a type in the INMV database [373]*. Patterns with the same number but different letters would be classified as the same type in the INMV database but have different repeat values. UNK = Unknown. These patterns can't be predicted due to a lack of certainty within specific loci.*

Table 2.3: Models selected by the IQ-Tree model finder for construction of each phylogenetic tree

Phylogeny Constructed	Model Selected[a]
A19 Specific Isolates	JC+ASC
A20 Specific Isolates	K2P+ASC
A21 Specific Isolates	K2P+ASC
A22 Specific Isolates	K2P+ASC
A23 Specific Isolates	K2P+ASC
A25 Specific Isolates	TIM2e+ASC
A26 Specific Isolates	K2P+ASC
A32 Specific Isolates	TIM2e+ASC
A34 Specific Isolates	TVMe+ASC
A36 Specific Isolates	K2P+ASC
A37 Specific Isolates	JC+ASC
AM1 Specific Isolates	K2P+ASC
AM2 Specific Isolates	TIM3e+ASC
AM3 Specific Isolates	TVMe+ASC+R2
All Isolates	TVMe+ASC+R2

[a]*Models selected based on the "Model Finder" function within IQ-TREE for each phylogeny.*

Table 2.4: ML-SSR and MIRU-VNTR patterns identified within select isolates extracted from multiple herds in Quebec and Ontario (n=139) as discussed in Chapter V.

Herd ID	Animal ID	Isolate Number	SSR1[a]	SSR2[a]	SSR3[a]	SSR4[a]	SSR5[a]	SSR6[a]	SSR7[a]	SSR8[a]	SSR9[a]	SSR10[a]	SSR11[a]	INMV SSR Type[b]	292[a]	X3[a]	25[a]	47[a]	3[a]	7[a]	10[a]	32[a]	INMV MIRU Type[b]
QC-2	A19	#159	18	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (6)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#164	18	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (6)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#167	18	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (6)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#160	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (7)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#162	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (7)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#168	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (7)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#166	14	9	5	5	5	5	5	5	5	5	5	MLSSR 7 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#161	15	9	5	5	5	5	5	5	5	5	5	MLSSR 7 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#165	12	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A19	#163	UNK	10	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#175	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#174	14	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#171	17	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (5)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#176	21	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (9)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#177	12	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#172	15	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (11)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#173	15	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (11)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#169	17	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (16)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A20	#170	UNK	14	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#179	16	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (4)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#180	20	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (8)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#183	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#184	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#187	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#185	18	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (5)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#182	15	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (12)	3	2	3	3	2	2	2	8	INMV 2

QC-2	A21	#186	17	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (16)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#188	19	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (20)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A21	#181	23	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (26)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#198	13	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#191	14	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#194	16	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (4)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#190	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#196	16	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#192	17	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (4)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#197	19	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (21)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#193	14	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (7)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#195	14	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (7)	3	2	3	3	2	2	2	8	INMV 2
QC-2	A22	#189	14	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (9)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#202	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#206	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#208	12	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#199	15	15	5	5	5	5	5	5	5	5	5	MLSSR 93 (13)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#204	15	16	5	5	5	5	5	5	5	5	5	MLSSR 93 (14)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#200	19	15	5	5	5	5	5	5	5	5	5	MLSSR 93 (23)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#201	12	18	5	5	5	5	5	5	5	5	5	MLSSR 93 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#205	13	15	5	5	5	5	5	5	5	5	5	MLSSR 93 (6)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#203	14	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (8)	3	2	3	3	2	2	2	8	INMV 2
QC-3	A23	#207	14	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (9)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#23	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#24	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#27	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#28	13	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#213	20	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (8)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#30	11	10	5	5	5	5	5	5	5	5	5	MLSSR 38	3	2	3	3	2	2	2	8	INMV 2

QC-1	A25	#26	15	9	5	5	5	5	5	5	5	5	5	MLSSR 7 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#25	16	9	5	5	5	5	5	5	5	5	5	MLSSR 7 (4)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#212	19	9	5	5	5	5	5	5	5	5	5	MLSSR 7 (5)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A25	#29	16	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (15)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#215	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#38	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#216	16	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (4)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#214	10	10	5	5	5	5	5	5	5	5	5	MLSSR 35	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#32	11	10	5	5	5	5	5	5	5	5	5	MLSSR 38	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#35	11	11	5	5	5	5	5	5	5	5	5	MLSSR 54	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#36	11	11	5	5	5	5	5	5	5	5	5	MLSSR 54	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#33	13	9	5	5	5	5	5	5	5	5	5	MLSSR 7 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#37	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
QC-1	A26	#34	13	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (4)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#113	18	14	5	5	5	5	6	5	5	5	5	N/P 11	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#119	18	16	5	5	5	5	6	5	5	5	5	N/P 12	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#120	18	18	5	5	5	5	6	5	5	5	5	N/P 13	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#116	13	14	5	5	5	5	6	5	5	5	5	N/P 2	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#121	14	12	5	5	5	5	6	5	5	5	5	N/P 3	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#122	15	12	5	5	5	5	6	5	5	5	5	N/P 4	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#117	15	13	5	5	5	5	6	5	5	5	5	N/P 5	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#114	17	12	5	5	5	5	6	5	5	5	5	N/P 8	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#118	17	15	5	5	5	5	6	5	5	5	5	N/P 9	3	2	3	3	2	2	2	8	INMV 2
ON-1	A32	#115	14	UNK	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#129	17	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (5)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#133	15	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (2)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#127	19	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (6)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#132	22	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (8)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#126	13	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (5)	3	2	3	3	2	2	2	8	INMV 2

ON-1	A34	#130	14	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (7)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#236	14	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (8)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#237	14	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (8)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#131	14	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (9)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A34	#128	19	13	5	5	5	5	6	5	5	5	5	N/P 15	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#247	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#249	24	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (10)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#255	18	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (6)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#252	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (7)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#254	20	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (8)	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#136	UNK	13	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#248	UNK	12	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#250	UNK	11	5	5	5	5	5	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#251	UNK	13	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A36	#253	UNK	12	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#262	13	10	5	5	5	5	6	5	5	5	5	N/P 1	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#259	18	10	5	5	5	5	6	5	5	5	5	N/P 10	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#263	19	10	5	5	5	5	6	5	5	5	5	N/P 14	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#265	19	10	5	5	5	5	6	5	5	5	5	N/P 14	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#137	20	10	5	5	5	5	6	5	5	5	5	N/P 16	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#264	16	10	5	5	5	5	6	5	5	5	5	N/P 6	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#261	17	10	5	5	5	5	6	5	5	5	5	N/P 7	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#256	UNK	11	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#258	UNK	10	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
ON-1	A37	#260	UNK	13	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#218	14	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#225	14	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (3)	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#217	18	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (6)	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#221	18	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (6)	3	2	3	3	2	2	2	8	INMV 2

QC-1	AM1	#219	11	10	5	5	5	5	5	5	5	5	5	MLSSR 38	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#220	11	11	5	5	5	5	5	5	5	5	5	MLSSR 54	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#224	11	11	5	5	5	5	5	5	5	5	5	MLSSR 54	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#222	12	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#223	12	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (1)	3	2	3	3	2	2	2	8	INMV 2
QC-1	AM1	#226	15	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (10)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#231	12	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (1)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#228	18	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (5)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#229	15	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (12)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#125	17	15	5	5	5	5	5	5	5	5	5	MLSSR 93 (18)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#234	18	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (19)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#233	19	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (21)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#124	19	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (22)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#232	20	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (24)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#230	20	14	5	5	5	5	5	5	5	5	5	MLSSR 93 (25)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM2	#123	14	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (7)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#239	19	10	5	5	5	5	5	5	5	5	5	MLSSR 10 (7)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#241	16	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (3)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#240	20	11	5	5	5	5	5	5	5	5	5	MLSSR 8 (7)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#135	17	12	5	5	5	5	5	5	5	5	5	MLSSR 93 (16)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#243	17	13	5	5	5	5	5	5	5	5	5	MLSSR 93 (17)	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#134	20	15	5	5	5	5	6	5	5	5	5	N/P 17	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#242	22	13	5	5	5	5	6	5	5	5	5	N/P 18	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#245	22	13	5	5	5	5	6	5	5	5	5	N/P 18	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#244	22	14	5	5	5	5	6	5	5	5	5	N/P 19	3	2	3	3	2	2	2	8	INMV 2
ON-1	AM3	#238	UNK	14	5	5	5	5	6	5	5	5	5	UNK	3	2	3	3	2	2	2	8	INMV 2
Discriminatory Index (DI) Value			0.9081	0.8051	0	0	0	0	0.3241	0	0	0	0	0.9839	0	0	0	0	0	0	0	0	0

[a] *Values listed as UNK are not included in DI calculations.*

[b]*N/P = New Patterns. Patterns are not recognized as a type in the INMV database. Patterns with the ML-SSR type but different numbers in brackets would be classified as the same type in the INMV database but have different repeat values. UNK = Unknown. These patterns can't be predicted due to a lack of certainty within specific loci.*

CHAPTER III

Methods for Detecting Mycobacterial Mixed Strain Infections – A Systematic Review

3.1 Abstract

Mixed strain infection (MSI) refers to the concurrent infection of a susceptible host with multiple strains of a single pathogenic species. Known to occur in humans and animals, MSIs deserve special consideration when studying transmission dynamics, evolution, and treatment of mycobacterial diseases, notably tuberculosis in humans and paratuberculosis (or Johne's Disease) in ruminants. Therefore, a systematic review was conducted to examine how MSIs are defined in the literature, how widespread the phenomenon is across the host species spectrum, and to document common methods used to detect such infections. The search strategy identified 121 articles reporting MSIs in both humans and animals, the majority (78.5%) of which involved members of the *Mycobacterium tuberculosis* complex (MTBC), while only a few (21.5%) examined non-tuberculous mycobacteria (NTM). In addition, MSIs exist across various host species, but most reports focused on humans due to the extensive amount of work done on tuberculosis. This study reviewed several strain typing methods that allowed for MSI detection and found a few that were commonly employed but were associated with specific challenges. This review also notes the need for standardization of some methods, as some highly discriminatory methods are well adapted to distinguish between the microevolution of one strain and concurrent infections with multiple strains. Further research is also warranted to examine the prevalence of NTM MSIs in both humans and animals. In addition, it is envisioned that the accurate identification and a better

understanding of the distribution of MSIs in the future will lead to important information on the epidemiology and pathophysiology of mycobacterial diseases.

3.2 Introduction

The genus *Mycobacterium* includes 195 different species with diverse growth characteristics [11,12] and host tropism [10]. Mycobacteria can be categorized based on whether a species can cause either tuberculosis (MTBC) [13], leprosy (including *Mycobacterium leprae* and *Mycobacterium lepromatosis*) [14], or neither disease. All other mycobacteria are commonly referred to as atypical mycobacteria [15], non-tuberculous mycobacteria (NTM) or mycobacteria other than *M. tuberculosis* (MOTT) [10]. Members of this genus can be further categorized based on their growth rates into rapid and slow growers, with the latter having prolonged doubling times, making it challenging to cultivate them [385].

Tuberculosis is caused by *M. tuberculosis* infecting the lungs of the host, though the pathogen can spread to other parts of the body [386]. Members of the MTBC such as *Mycobacterium africanum* also cause tuberculosis in humans [387], while non-human host tropism is reported for other bacteria from the group. For example, *Mycobacterium bovis* causes bovine tuberculosis [388–390], *Mycobacterium caprae* can infect a variety of wild and domesticated animals, and *Mycobacterium pinnipedii* causes tuberculosis in pinniped species [391]. Tuberculosis is an ancient disease afflicting humans, and while *M. tuberculosis* has been studied for over a century, the disease remains a significant cause of global morbidity and mortality [392]. One reason why tuberculosis remains problematic is due to the complex interaction between MTBC members and their hosts, many aspects of

which are still not fully understood. In addition, the emergence and spread of drug-resistant forms of *M. tuberculosis* further exacerbate the situation, leaving few effective treatment options in some cases [392].

The NTM group is comprised of over 150 different species, including several pathogens from the *Mycobacterium avium* and *Mycobacterium abscessus* complexes [393]. Members of the MAC are commonly found in the environment and cause opportunistic infections [28,30,394,395], especially in immunocompromised individuals such as those suffering from acquired immunodeficiency syndrome (AIDS) [396–398]. Furthermore, *M. avium* includes several subspecies, which may infect organs other than the lungs, including MAP.

The *M. abscessus* complex includes three fast-growing subspecies (*abscessus*, *massiliense* and *bolletii*), which are highly resistant to many antibiotics and cause a wide range of human infections [399–402]. Another NTM of significance is *M. genavense*, an opportunistic pathogen that often causes disease in immunocompromised patients and has also been found to infect various domestic companion animals [403–409]. The NTM discussed above are just a few of many that are of concern to human and animal health [410–414], demonstrating the propensity of members from this group to cause diverse diseases if given the opportunity.

The progression and outcomes of an infection are dependent on many factors, including the resident host-microbiome and the presence of other pathogens, sometimes from the same genus **(Figure 1.2)** [291–293]. Mixed-species infections refer to the phenomenon where different species belonging to the same genus concurrently infect a

single host. Another important factor to consider is the potential for genetically distinct strains (or isolates) of the same pathogenic species to infect a single host at any given time, which is sometimes referred to as a polyclonal infection [294–296]. This situation can potentially arise if an isolate undergoes intra-host evolution (also referred to as microevolution) following infection, leading to minor genetic differences in the resulting progeny [297–299]. Another mechanism leading to polyclonal infections involved concomitant or sequential infection by genetically distinct strains [294,304,305]. Therefore, by examining MSIs and their transmission, successful treatment methods can be devised, and essential information might also be gained for use in future vaccine development endeavours.

3.3 Objectives

The main objective of this study was to conduct a systematic review to gain a better understanding of MSIs across the genus *Mycobacterium* and the methods used to detect them. Typically, the detection of such infections is challenging due to the lack of distinct intraspecies markers that allow for the discrimination between isolates/strains. Despite this, MSIs in mycobacteria were found using a variety of strain typing methods, each with a different level of discriminatory ability and ease of use, with different methods focusing on specific aspects of the *Mycobacterium* genome. Mycobacterial strain discrimination is made possible by RFLP analysis in species-specific insertion sequences IS-elements such as *IS6110* (associated with the MTBC, with some exceptions) [415,416], and *IS1245* or *IS1311* (both associated with the MAC) [55,417,418]. Another general method used to discriminate between strains exploits the nucleotide sequences present in VNTRs,

which are dispersed throughout mycobacterial genomes. By examining differences in the number of nucleotide repeats present at distinct loci, individual strains can be typed. Different mycobacteria harbour a variety of VNTR loci, though depending on the species and loci examined, they may instead be referred to as multi-locus variable-number tandem repeats [419–421], MIRU-VNTR [422,423] or SSRs [286,315]. Analysis of the entire genome at the individual nucleotide level using methods based on WGS also allows for examination of strain diversity but at a resolution unmatched by RFLP or VNTR-based methods. By using WGS, strains can be typed and compared without focusing on a given set of loci allowing for more accurate detection of MSIs, re-infections, and relapses [417,424–426]. Heterogeneous (also referred to as heterozygous) SNPs are predominantly used in strain comparisons, and the presence of many different SNPs in isolates from a single sample is suggestive of MSIs [427].

Another intention of this review was to help clarify what constitutes a true MSI as compared to similar events such as re-infection, relapses, and microevolution. While polyclonal infection may refer to microevolution, some studies have also used the term to describe infections that fit the criteria of an MSI [316–319,428]. Due to this lack of consensus regarding the terminology used in the literature and to be consistent in this review, definitions for a selection of terms were developed to describe different events **(Table 3.1)**. For this review, MSIs refer to an infection where multiple unrelated strains, which did not evolve from an initial infecting strain, are present within a single host at the same time.

3.4 Results and Discussion

An initial screen of the literature yielded 14,952 records, and after the removal of duplicate and non-relevant entries based on abstracts and content, 253 articles were retained for full-text review **(Figure 2.1)**. Examination of these articles resulted in the further exclusion of 156 entries for various reasons as described in **Figure 2.1**, leaving 97 reports for inclusion in the review. Additionally, 24 other relevant studies cited in the 87 reports were also included and presented in the modified Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method, to adhere to the systematic review format **(Figure 2.1)**.

Data was extracted from all 121 selected articles, and general information, including the publication year, primary author, study location, and bacterial and host species involved, was recorded **(Appendix Table A2)**. The number of samples/isolates, if they were derived from clinical specimens or cultures, the prevalence of MSIs reported in each study and the human immunodeficiency virus (HIV) status of human subjects were also noted when possible. The studies were allocated into two separate groups based on MSI reports in humans and animals, respectively **(Table 3.2)**.

3.4.1 *Mycobacterium tuberculosis* Complex

One of the earlier methods developed to discriminate between strains of *M. tuberculosis* involved the use of mycobacteriophages **(Table 3.3)**, which specifically lyse certain strains, leading to the formation of plaques on solid agar plates [429]. Phage typing has also been used to identify cross-contamination, transmission dynamics and MSIs based on the sensitivities of *M. tuberculosis* isolates to a panel of selective phages [430–433],

but more modern methods are faster and offer better discriminatory power [433]. The primary systematic review workflow did not find any reports on the identification of MSIs using phage typing, but secondary searches found two such studies. Mankiewicz & Liivak [432] sampled 233 patients, of which 33 (14.2%) showed evidence of MSIs due to the presence of multiple *M. tuberculosis* phage types in a single culture. In the other study, Bates et al.[430] analyzed samples from 87 different patients and identified three (3.4%) as having mixed phage-typing profiles. While the possibility that the presence of multiple *M. tuberculosis* phage types within the same patient could indicate an MSI, the limited discriminatory power of the method prevents definitive confirmation and cannot completely rule out intra-host microevolution.

Until recently, RFLP based on the insertion sequence *IS6110* was the standard method used for comparing the genetic relatedness of *M. tuberculosis* isolates [434,435]. *IS6110* (sometimes also referred to as *IS986*) belongs to the *IS3* family, members of which are only present in the MTBC [436–438]. *M. tuberculosis* and *M. bovis* strains can contain 0-25 and 0-3 copies each of *IS6110*, respectively [439–447]. Variations in the copy number of *IS6110* elements within different *M. tuberculosis* strains make it an attractive target for epidemio-typing isolates containing multiple copies of the insertion sequence, but not in low copy number strains. Therefore, *IS6110*-typing has led to the development of extensively used standardized protocols **(Table 3.3)** [448].

In total, 26 (21.5%) of the publications reported herein include the use of *IS6110*-typing methods for detecting MSIs involving MTBC members [301,303,316,449–471], 11 of which exclusively used *IS6110*-RFLP **(Appendix Table A2),** and one study used *IS6110*-

inverse PCR [463]. All of these studies employed cultures in their analysis, and *IS6110*-inverse PCR was primarily used as a tool to identify strains belonging to the Beijing evolutionary lineage [463,472]. The first report of MSI detection using *IS6110*-RFLP was by Yeh et al. [471], who noticed multiple bands of varying intensities in a sample from a patient due to the presence of two separate strains. Another study found an MSI with two drug-susceptible isolates in an immunosuppressed patient [465], and similar infections with drug-susceptible and drug-resistant *M. tuberculosis* have also been detected by *IS6110*-RFLP in separate reports [452,473]. Overall, MSIs were identified in 0.4-100% of cases using *IS6110*-RFLP in these studies **(Appendix Table A2)**, where most reports with 100% MSI detection rates involving a single patient [451,465,471]. In addition, a recent study detected MSIs in 3 out of 17 samples from patients using two probes for *IS6110*-RFLP instead of the conventionally used single probe [450].

Spacer oligotyping (spoligotyping) is a commonly used method for the simultaneous detection and identification of MTBC members **(Table 3.3)** [474]. Spoligotyping exploits the nucleotide sequence diversity of clustered regularly interspaced short palindromic repeats (CRISPRs) [475], which are present in many bacteria and archaea [476]. The chromosomal locus specifically used in this assay is known as direct repeat (DR) in mycobacteria [477]. Spoligotyping is traditionally performed by amplifying the entire DR region using PCR with a pair of oligonucleotide primers, one of which is labelled with biotin to aid in the detection of products by hybridization. Membranes containing a unique set of 43 covalently bound synthetic oligonucleotide spacer sequences derived from *M. tuberculosis* and *M. bovis* Bacille Calmette-Guérin (BCG) are

used in the hybridization [474], and can differentiate between MTBC isolates based on the presence or absence of spacers.

Spoligotyping has been widely used in epidemiological studies to investigate the cause of recurrent tuberculosis (defined as endogenous reactivation of an initially infecting strain or exogenous reinfection with a different strain) [301,478–480], tracking epidemics [481–483], and investigating laboratory cross-contamination [484]. Twenty (16.5%) publications reported using spoligotyping as one of the methods for detecting MSIs, with a majority employing one [419,455,460,465,485–492], or two [459,462,463,493,494] additional methods **(Appendix Table A2)**. Only three reports used spoligotyping as the sole genotyping method and reported MSIs at frequencies ranging from 11.8-57.1% [478,495,496].

One significant limitation of spoligotyping is that it can underestimate MSIs, as hybridization signals from multiple strains in a sample can overlap and appear as a single pattern [496]. For this reason, when spoligotyping is used to investigate MSIs, subculturing is usually performed to obtain single isolated colonies for testing [459,462,467]. The ability to detect MSIs in the latter case is dependent upon the proportion of different strains in the initial sample and the number of colonies picked for analysis. To help resolve this problem, Lazzarini et al.[497] developed a computational method which can predict if individual spoligotypes contain signatures from more than one of four major global lineages, which would indicate an MSI. In most cases, a secondary typing method like MIRU-VNTR, *IS6110*-RFLP, or WGS may be required to verify results that may appear to contain a single spoligotyping pattern. It is worth noting that although spoligotyping can be applied directly to clinical specimens, all the studies reported herein used cultures,

possible due to the requirement of purified DNA for other methods employed by the authors [467,473,498]. Warren et al.[499] used a combination of lineage-specific PCR and spoligotyping to distinguish between *M. tuberculosis* strains belonging to the Beijing- and non-Beijing evolutionary lineages. They detected MSIs in 19% of samples from patients associated with retreatment in their study. In addition, WGS was used as a secondary technique in three spoligotyping-based studies [426,485,494], one of which focused on animals. In their work, Silva-Pereira et al.[494] detected an *M. pinnipedii* MSI in a South American sea lion (*Otaria flavescens*) using WGS, which was not suggested by *in-silico* spoligotyping or MIRU-VNTR initially. The above-mentioned studies demonstrate the importance of using more discriminatory methods along with traditional screening techniques to ensure the detection of different strains that might be present in a single sample.

VNTRs are short DNA sequences, which are dispersed throughout the genomes of many bacterial and eukaryotic species [500,501]. They vary in repeat unit length and repeat number depending on the specific organism and locus being analyzed [502]. Since the repeat unit length at specific VNTR loci is known for each species, determining the number of repeats present at the respective loci can be used to discriminate between strains **(Table 3.3)**. The use of VNTRs for typing *M. tuberculosis* strains was first reported in 1998 [503], and since then the discriminatory power has been improved by using combinations of mycobacteria-interspersed repetitive units (MIRUs) located at different loci throughout the genome [504–507]. Initially, a 12-locus MIRU-VNTR method was widely used [422], but the method has some limitations in its ability to discriminate between unrelated isolates [423].

To overcome this problem, the stability and resolution power of 29 MIRU-VNTR loci was evaluated using predominant *M. tuberculosis* lineages, resulting in the standardization of 24 MIRU-VNTR loci for high-resolution epidemiological studies [423]. In addition, the 15 most discriminatory loci of the 24 were selected for use in routine epidemiological investigations involving *M. tuberculosis* [423].

To generate a MIRU-VNTR profile, several genomic regions known to contain VNTRs are amplified by PCR using specific primer pairs either individually (simplex) or in multiples (multiplex). In this way, the number of repeats at each VNTR locus can be determined using different DNA sizing techniques for comparing isolates. It was found that 50 studies used VNTR-typing to detect *M. tuberculosis* MSIs, where 28 studies used it was the sole discriminatory method for this purpose [300,307,309,310,317–320,419,449,453,461–464,467,473,486–494,508–531]. In addition, 24-locus MIRU-VNTR was the most commonly used method for detecting *M. tuberculosis* MSIs, and nearly all (47/50, 94.0%) of the reports used some form of culturing for the analysis. Comparatively, MIRU-VNTR detected more MSIs than any other method based on the current review **(Table 3.2)**. One reason for this could be the use of PCR amplification during MIRU-VNTR, which increases the sensitivity and detection power of the method [468], especially in instances where different strains are not proportionally present in a single sample.

Among other PCR-based methods used to detect MSIs, the majority were focused on differentiating between *M. tuberculosis* strains belonging to the Beijing and non-Beijing evolutionary lineages [301,304,499,532,533]. For example, Warren et al.[468] detected MSIs in 35 (18.8%) of the 186 sputum cultures from tuberculosis patients tested during an

epidemiologic study in South Africa. They reported that MSIs were more often associated with retreatment (23%) vs. new cases (17%), and the sensitivity and specificity of their method were comparable to *IS6110*-RFLP and spoligotyping [499]. Using the same method, van Rie et al.[301] also detected one case of MSI with drug-susceptible and drug-resistant *M. tuberculosis* isolates. While the method developed by Warren et al.[499] used simplex PCR, Huang et al.[304] utilized multiplex PCR to detect MSIs caused by Beijing and non-Beijing lineage strains in 11.3% of the 185 sputum samples from patients without any prior history of tuberculosis treatment. Another group used quantitative PCR to detect MSIs based on the presence of both Beijing and non-Beijing lineages in 3% of tuberculosis patients <25 years of age using clinical specimens and cultures [533], whereas *M. tuberculosis* isolates belonging to the two lineages were also detected together in 14.7% of cases by Mustafa et al.[532]. In contrast, a Latin American and Mediterranean (LAM) and non-LAM lineage-based PCR found MSIs in 4 out of 160 (2.5%) culture-positive sputa analyzed [534], suggesting that such methods are useful in identifying MSIs under settings where the occurrence of *M. tuberculosis* isolates from mixed lineages is high. Therefore, by using an in-house PCR to identify isolates from the Beijing, Haarlem, S-family, and LAM evolutionary lineages, Hanekom et al.[535] were able to detect MSIs in 31 (15%) of the 206 samples analyzed in their study.

M. tuberculosis MSIs have also been reported in patients with discordant drug susceptibility profiles on more than one occasion. Isolates from 10 such tuberculosis patients out of 89 (11.2%) were confirmed to have MSIs using Beijing lineage-specific PCR [536] and 16-locus MIRU-VNTR [525]. Other techniques such as linker PCR [537] and *gyrA*

PCR/sequencing [538,539] have also detected *M. tuberculosis* MSIs in cases involving discordant drug susceptibility profiles and also in archeological samples. In addition, two studies employed double repetitive PCR [303,540] based on the *IS6110* and a GC-rich repetitive sequence described by Friedman et al.[541]. Baldeviano-Vidalón et al.[273] also observed multiple discrepancies in drug susceptibility testing results among follow-up samples from patients, which were attributed to MSIs based on *IS6110*-RFLP analysis. Such studies emphasize the importance of considering MSIs during tuberculosis drug susceptibility testing and while devising appropriate treatment regimens.

Methods based on WGS provide the ability to examine strain diversity at very high resolution, which cannot be achieved by other techniques such as RFLP or MIRU-VNTR. Heterogeneous SNPs are predominantly used for strain discrimination, and the presence of many different SNPs in isolates from a single sample is suggestive of MSIs [427]. While the concept is simple, where a SNP is confirmed within several sequencing reads used to assemble the locus being examined **(Table 3.3)**, the technical criteria used for identifying bona fide SNPs varies between studies. Factors that can affect SNP detection include the quality and depth of sequencing, experimental design, sample preparation and pathogen species, to name a few [542]. Some reports require that the frequency of the alternative base at a specific locus be found from anywhere between 5% and 30% or more of the reads for SNP calling [543–545], or even just two reads in some deep sequencing studies [546]. Some studies also include threshold nucleotide base quality scores to minimize artifacts [543,546,547]. Additionally, several loci with heterogeneous bases must be identified between isolates to qualify them as MSIs. However, the minimum number of

SNPs used to qualify the presence of an MSI using isolates also varies, as anywhere over 16 to 80 have been used for the purpose depending on the sequencing technology [543,546]. In addition, WGS analysis cannot be performed directly on clinical samples in most cases because the genetic complexity of the sample limits the confidence at which SNPs are called. Therefore, WGS often requires axenic cultures for strain typing and MSI identification [289,548,549].

Of the 97 studies identified in Covidence, eight (8.2%) primarily used WGS to detect MSIs in humans [426,485,543,544,546,550–552], while three others (3.1%) used the method to detect MSIs in various animal species [289,494,553], only one of which focused on an MTBC bacteria [494]. Secondary searches found four more human studies on WGS and mycobacterial MSIs [547,554–556], in addition to reports where the method was used to confirm such infection that was initially identified using other means **(Appendix Table A2)**. For example, six isolates from 47 paired patient samples taken before and after treatment during the REMoxTB clinical trial were initially classified as relapses or re-infections by MIRU-VNTR but were later determined to be MSIs by WGS analysis [543]. A follow-up study re-examined the same data using QuantTB [550], a tool developed to identify MSIs through the iterative comparison of SNPs and suggested that only four of the six MSI cases could be classified as such. O'Donnell et al.[551] used isolates from a patient where drug susceptibility testing alluded toward an MSI involving susceptible and resistant *M. tuberculosis*, which was confirmed using WGS. They showed that the patient was initially infected with a drug-susceptible strain followed by an extensively drug-resistant *M. tuberculosis* isolate, which was selected during antibiotic therapy. The high

resolution of WGS underscores its importance in strain typing for devising individualized treatments for tuberculosis infections, although its widespread use may be limited in many high tuberculosis burden settings due to insufficient resources or technical capabilities. The ability to use traditional culture-based drug susceptibility testing methods has limitations for many slow-growing pathogenic mycobacteria, but genomics-based technologies are more rapid and allow for the detection of resistance based on the presence of conferring mutations [557]. While PCR methods targeting specific genes can detect important drug resistance mutations for early diagnosis, WGS can additionally infer potential resistance, allowing for individualized treatment regimens [557].

The use of WGS also enhances the detection of potential MSIs. For example, while examining pre- and post-treatment isolates from a tuberculosis patient, Witney et al. [395] noticed 57 SNP differences between the two of them, indicating a re-infection by a second strain. More detailed analysis of the WGS data indicated a potential MSI in the pre-treatment sample at a 3:1 genotypic ratio by two strains, where the minor genotype was closely related to the post-treatment isolate. This suggested that recurrent disease was caused by a relapse, where one of the two strains from the MSI was eliminated during the initial course of therapy. Interestingly, 24-locus MIRU-VNTR typing did not detect the genotype of the post-treatment isolate in the pre-treatment sample, which is intriguing, as the method was previously shown to detect MSIs in proportions as low as 1:99 [460]. Such reports suggest that many MSIs might have gone undetected due to technical limitations and could have potentially affected disease outcomes.

Although more tuberculosis MSIs are now being reported, evidence from archaeological studies indicates that the phenomenon has been around for a long time. Through metagenomics analysis of ancient DNA, one study identified an MSI (difference of 398 SNPs) within an eighteenth-century Hungarian mummy [554]. A follow-up study by a related group reported five MSIs in eight mummified bodies examined from the archaeological site using similar metagenomics-based methods [556]. In addition, a separate study employed *gyrA* PCR to successfully identify an MSI using ancient DNA from a variety of archaeological samples across Britain and France [538]. The detection of MSIs using ancient DNA and complex samples is intriguing. Such reports also provide precedence for using metagenomics and other technologies to examine the prevalence and impact of such infections in future prospective and retrospective studies.

3.4.2 Non-tuberculous and Other Mycobacteria

This review showed that MSIs involving NTM have not been investigated to the same extent as compared to those caused by *M. tuberculosis* **(Table 3.2)**. Amongst the 121 mycobacterial studies identified, 26 (21.5%) examined NTM, of which 17 (14.0%) and eight (6.6%) found MSIs in humans and animals, respectively, whereas one report identified MSIs in both human [428,558–573] and animal [289,315,488,553,574–578] populations simultaneously [579]. Many early studies used pulse field gel electrophoresis (PFGE) to identify MSIs in 14.3-100% of patients infected with MAC bacteria [558,565,571–573]. PFGE was also used as the sole method to identify MSIs associated with other NTMs. For example, Legrand et al.[532] found 33.3% *M. simiiae* MSI prevalence (3 of 9 hosts) in their study involving AIDS patients. In addition, PFGE was used in combination with other

methods in studies on NTM MSIs. A report by Picardeau et al.[568] detected 3 *M. avium* MSIs while examining 93 samples from AIDS patients using simple double repetitive element PCR (*Ma*DRE-PCR), which amplifies a region of the *M. avium* chromosome between *IS1245* and *IS1311* [580]. They confirmed these results using *IS1245*-RFLP, which showed the presence of multiple low-intensity bands in the same sample. PFGE was also able to pick up multiple banding patterns, including some samples from other patients [568], but the criteria for attributing them to different strains (≥3 differences) was not surpassed [581]. A similar study on 31 AIDS patients from the Caribbean islands initially identified three potential polyclonal *M. avium* infections based on *IS1245*-RFLP, but PFGE analysis showed that two isolates had identical banding patterns [564]. Therefore, results from PFGE analysis do not corroborate those obtained by other methods on multiple occasions, which may allude to differences in their discriminatory powers.

IS1245 and *IS1311* have also been used as general fingerprinting targets to identify an MSI (25.0%) involving MAC members from 25 patients [559]. In a separate study, Oliveira et al.[566] used RFLP analysis of *hsp65* PCR products, *IS1245* and *IS1311*, respectively, to detect an *M. avium* MSI. Though the presence of *IS1245*-RFLP may be supplemented using additional methods, studies have shown it alone is capable of identifying 2.6-100% of NTM MSIs in samples from human subjects, including those with positive HIV-AIDS status **(Appendix Table A2)** [560,567–569,579]. In addition, a study examined 41 samples from 14 AIDS patients using *Ma*DRE-PCR, which found four cases with multiple banding patterns, though re-evaluation using *IS1245*-RFLP only confirmed two as *M. avium* MSIs [570]. Based on the reports mentioned above, the feasibility of using

IS1245-RFLP as a screening method for detecting NTM MSIs warrants further evaluation.

Kimizuka et al. [531] have also used a variety of VNTR loci, including 16 *M. avium* tandem repeats (MATR) and five Higashi Nagoya tandem repeats (HNTR) to identify *M. avium* MSIs in samples from nine patients (13.8%) out of 65 examined. Another study examined samples from 120 patients with pulmonary MAC infections (94 *M. avium* and 26 *M. intracellulare*) [428]. MIRU-VNTR analysis using 15 loci for *M. avium* and 16 loci for *M. intracellulare* successfully identified 20 and seven cases of *M. avium* and *M. intracellulare* MSIs, respectively. Other methods that have detected NTM MSIs in humans include random amplified polymorphic DNA (RAPD) analysis [582]. Using RAPD-based methods, García-Pedrazuela et al.[561] identified MSIs involving several species, including *M. abscessus*, *Mycobacterium chelonae*, *Mycobacterium fortuitum* and *Mycobacterium mucogenicum*, while examining 64 isolates from Spanish patients.

Systematic screening through Covidence also identified some reports on NTM MSIs in animals, including a few involving MAP. Gerritsmann et al.[575] examined 39 MAP-containing samples from a variety of wild and domestic ruminant species, five of which were classified as MSIs using an 8-locus MIRU-VNTR method **(Appendix Table A2)**. In another study, Gioffré et al.[576] also used the 8-locus MIRU-VNTR method to identify a MAP MSI from a group of 97 cattle, sheep, and goats based on differences in two loci. MAP MSIs have also been identified using ML-SSR typing [315], which examines small repeat sequences that vary between isolates and allow for their discrimination [286]. In addition, the use of DNA detection and sizing techniques such as fragment analysis of

labelled PCR products further improves the resolution of SSR typing [315,583]. Using this strategy, Podder et al. [315] identified MAP MSIs in all 18 animals from their study, which was subsequently confirmed by WGS analysis using some of their isolates [289]. MAP strains from the same animal had significantly different SNPs at high frequencies, which ruled out microevolution based on evolutionary rates [53].

In addition to MAP, there were reports on MSIs present within other NTM in animals. *M. avium* subsp. *avium* MSIs were described in domestic chickens using *IS901*-RFLP, where multiple banding patterns were detected in 7 of the 16 (43.8%) tissue samples tested [578]. 8-locus VNTR was used to detect *M. avium* subsp. *hominissuis* MSIs in pigs using isolates from multiple organs of a single animal [577]. Another study that previously reported a single *M. avium* subsp. *hominissuis* MSI in a human (out of 26 patients, 3.8%) using *IS1245*-RFLP also detected similar infections in 33 pigs (13.5%) [579]. An additional study by Pfeiffer et al. [522] used WGS to detect non-MAP MSIs from 113 birds, reporting 12 cases of MSIs (2 involving *M. avium* and 10 involving *M. genavense*) based on differences in at least 12 SNPs, as suggested by Walker et al. [584]. Therefore, NTM MSIs seem to occur in a variety of animal species, but studies examining their prevalence and impact are few and far between.

This review was unable to reveal the occurrence of MSIs within leprosy-causing mycobacteria. However, using artificially co-infected armadillos, Shin et al. [527] demonstrated that distinct strains of *M. leprae* could simultaneously exist within the same host. It was noted that the *in-vivo* growth rate of the non-armadillo strain was significantly higher in the absence of competing strains, suggesting that pathological

variations exist between the different strain types. Due to the challenges associated with culturing *M. leprae* [585], the same experimental approaches used to study MTBC and NTM members might not be feasible, thereby limiting the potential for detecting *M. leprae* MSIs. Therefore, with the advent of more sensitive and powerful discriminatory methods, the use of culture-independent techniques such as metagenomics may help to shed light on MSIs involving *M. leprae* in the future.

3.4.3 Challenges Faced in Defining MSIs

Many questions are raised regarding the use of standard strain typing methods for delineating true MSIs from microevolution. When using *IS6110*-RFLP and MIRU-VNTR on *M. tuberculosis* isolates derived from the same host, MSIs are identified based on differences in specific DNA/PCR fragment profiles **(Table 3.4)**. *IS6110*-RFLP profiles that differ by at least 2-3 bands indicate the occurrence of concomitant infection by distinct strains [467,468,586]. Conversely, minor variations, i.e., anything <2 discriminatory bands, are considered microevolution. In the case of MIRU-VNTR, MSIs are predicted based on heterogeneity at two or more loci [467,468,511,587], whereas differences in a single locus suggest microevolution [467,468,587]. Therefore, allelic diversity at more than one locus is a criterion for differentiating microevolution from true MSI when using MIRU-VNTR.

Defining MSIs using WGS-based methods also has inherent challenges. However, the high discriminatory power of WGS allows for more definitive explanations for tuberculosis recurrence post-treatment [543–545]. The classification of infections based on SNP differences between isolates from a single patient varies, where 5-10 suggest the occurrence of a relapse and anything over 100, a re-infection [543–545] **(Table 3.4)**. These

differences in SNP numbers are set by taking into account the calculated evolutionary rate of *M. tuberculosis*, which ranges from 0.1-0.5 SNPs per genome per year, and limits the number of SNPs that can accumulate within strains in a defined period of time [53,290,313,584,588,589]. Varying SNP thresholds have been used to define MSIs, where Bryant et al. [543] called them such if isolates derived from the same patient had 80 or more different SNPs based on manual inspection. In comparison, Guerra-Assunção et al. [514] used a threshold of 140 heterogeneous SNPs to define MSIs based on an empirical cut-off generated during their data analysis. Furthermore, Dippenaar et al. [544] did not define a specific threshold, as the numbers of heterogeneous SNPs in their study were either limited (1-2, classified as microevolution) or rather numerous (757-883, classified as MSI). These differences in SNP numbers used to identify *M. tuberculosis* MSIs show that the field requires some form of standardization so that results can be compared between studies. The lack of a clear definition for MSIs using WGS may hinder the identification of some reports for inclusion in this review. Although this is unlikely as only a few studies have used WGS technology for such purposes to date, and most of them were manually screened for relevance.

Considering that the criteria necessary for identifying different types of *M. tuberculosis* polyclonal infections using *IS6110*-RFLP and MIRU-VNTR are well-defined, several studies were also found in the current review that reported on microevolution [308,310,590,591]. The classification was not evident in some other reports that used *IS6110*-RFLP or MIRU-VNTR, due to results potentially indicating either an MSI or a microevolution event depending on the applied criteria [592,593], or due to WGS not having

a well-defined SNP threshold [594]. While microevolution may initially appear to have little significance when compared to MSIs in terms of virulence and pathogenicity, one study found that the presence of highly evolvable repeats near genes (VNTR52, QUB26, or MIRU10/27) can influence gene expression in different *M. tuberculosis* isolates [595]. In addition, microevolution has obvious implications for the emergence of drug resistance due to selective pressures applied during treatment regimens [596].

Microevolution-derived infections may not be as distinct as MSIs, but different clonal variants can still spread in unique patterns. While a strain undergoes a genetic drift, some of the progeny might spread both within and between hosts. For example, WGS identified separate *M. tuberculosis* clonal variants derived from within host microevolution at both respiratory and extrapulmonary sites in a patient [597,598]. A study by Buff et al. [599] examined several tuberculosis cases where *IS6110*-RLFP and spoligotyping of isolates from community transmission events showed identical patterns but exhibited variations at a single locus during 12-locus MIRU-VNTR analysis. It was shown that *M. tuberculosis* isolates from the source patient also displayed varying MIRU-VNTR profiles that matched the secondary patients, suggesting that different clonal variants were transmitted individually.

In addition to the MSIs described in this review, there is also evidence for the microevolution of mycobacterial pathogens in animals. For example, microevolution events and MSIs involving *M. avium* subsp. *hominissuis* in bongo antelopes were identified using 8-locus MIRU-VNTR and *IS1245*-RFLP [600]. Another study used 8-locus MIRU-VNTR to examine *M. caprae* from 55 different animal hosts (including goats,

cattle, sheep, and wild boar) in Portugal and identified microevolution-derived infections in 12 of them [601]. Therefore, mycobacterial microevolution also occurs in animals, which is expected given the current state of knowledge regarding such infections in humans.

3.5 Conclusion

As summarized in this systematic review, a variety of terminologies have been used for describing MSIs, some of which overlap with microevolution. Findings show that MSIs exist among many different mycobacteria, although the majority of studies have been conducted in humans and predominantly focus on *M. tuberculosis*. In addition, most studies used VNTR-based methods, though more recent reports involved WGS. This change in methodology may represent an overall shift as newer technologies are developed and used more widely for strain typing. With methods suitable for large-scale screening of genetic variations by the massively parallel sequencing approaches, the potential for the accurate identification of MSIs is now accessible at a lower cost. Methods based on WGS offer an unprecedented resolution but appear to lack uniform SNP thresholds [543–545], sometimes making it ambiguous to differentiate between a true MSI or microevolution event. Therefore, there is scope for further standardizing WGS criteria for discriminating MSIs from genomic drift or technical aberrations, but as of now a universal definition for calling mycobacterial MSIs using the technology is lacking. Consideration of both microevolution and MSI has potential implications for developing personalized medical treatments for diseases such as tuberculosis. A select number of publications examined the historical context of *M. tuberculosis* using ancient DNA [538,554,556], suggesting that MSIs have likely been around for a long time. Furthermore,

reports on discrepant drug susceptibilities between one or more *M. tuberculosis* isolates from the same host during a single disease episode underscores the importance of considering MSIs when managing mycobacterial infections. Reports have shown that tuberculosis patients with MSIs associated with heteroresistance are at a higher risk of poor treatment outcomes [300,308,318,530]. Therefore, additional strain typing is recommended under certain incidences where heteroresistance is detected, to determine if it is caused due to MSIs or microevolution. It was also noted that studies on NTM MSIs in humans and animals are limited, but such infections are found across many species. Given the importance of NTM in causing opportunistic and nosocomial infections in humans and diseases in farmed animals, the prevalence and impact of MSIs caused by this large and important group of mycobacteria warrant further investigation.

3.6 Figures and Tables

3.6.1 Tables

Table 3.1: Glossary of terms developed for the systematic review.

Term[a]	Description[b]
Sample/specimen	A sputum, blood, feces, or otherwise uncultured biological sample from a host/patient, which is directly used for strain-typing analysis or is subsequently used for culturing.
Culture	Refers to the amplification of bacteria following growth on solid or in liquid media using a sample/specimen as an inoculant. Does not explicitly imply a pure or axenic culture.
Isolate (strain, if characterized)	Refers to a single colony from an agar plate or an axenic bacterial culture derived from a sample/specimen.
Microevolution	A co-infection that involves multiple strains/sub-strains in a single host that evolved from a single strain that caused the initial infection.
Mixed genotype infection (MGI)/polyclonal infection	A co-infection caused by two or more strains of the same species in a single host. Encompasses both microevolution and mixed-strain infections.
Mixed strain infection (MSI)	A co-infection caused by phylogenetically distinct strains of the same species in a single host (contrary to microevolution). Sometimes also referred to simply as mixed infection in review.
Re-infection	Recurrent disease caused by a strain that is unrelated to the one that caused the initial infection.
Relapse	Recurrent disease due to the same strain that caused the initial infection.

Table 3.2: General characteristics of reviewed mycobacterial studies involving MSIs (n=121).

Attribute	Human (%)[a]	Animal (%)[a]
Mycobacterial group responsible for MSIs		
M. tuberculosis complex (MTBC)	87 (71.9%)	8 (6.6%)
Non-tuberculous Mycobacteria (NTM)[b]	18 (14.9%)	9 (7.4%)
Geographic location[c]		
Africa[b]	28 (23.1%)	4 (3.3%)
Asia[b]	31 (25.6%)	0 (0.0%)
Europe[b]	20 (16.5%)	8 (6.6%)
North America	15 (12.4%)	3 (2.5%)
Oceania	1 (0.8%)	0 (0.0%)
South America[b]	11 (9.1%)	3 (2.5%)
Publication year (decade)		
Pre-2000	12 (9.9%)	0 (0.0%)
2000-2009[b]	32 (26.4%)	4 (3.3%)
2010-2019	54 (44.6%)	13 (10.7%)
2020	7 (5.8%)	1 (0.8%)
MSI detection method[d]		
Insertion Sequence Based RFLP[b]	34 (28.1 %)	3 (2.5%)
Variable Number Tandem Repeat	44 (36.4%)	12 (9.9%)
Whole Genome Sequencing	12 (9.9%)	3 (2.5%)
Spoligotyping	14 (11.6%)	6 (5.0%)
Region Specific PCR[b]	16 (13.2%)	0 (0.0%)
Other[e]	11 (9.1%)	0 (0.0%)

[a]Studies were grouped based on multiple criteria (host, geographical location, publication date and methods used). Totals for some sections may not add up to 100% as certain studies met multiple criteria and were therefore counted more than once.
[b]This subsection contains studies counted under more than one criterion.
[c]Excluding one study [546], which used information on isolates from a global database.
*[d]Not counting methods that were unable to identify MSIs in the respective studies, Methods are briefly described in **Table 3.3** and details are included in **Appendix Table A2.***
*[e]Includes methods such as phage typing, PFGE, etc. Details are provided in **Appendix Table A2**.*

Table 3.3: Overview of methods that were used to detect mycobacterial MSIs.

Method[a]	Target[b]	Output or readout[c]	Sample type[d]
Insertion sequence based RFLP	Species-specific insertion sequences	Differences in restriction enzyme band hybridization patterns	Culture
Variable number tandem repeat	Species-specific tandem repeat loci	Differences in copy numbers based on PCR	Specimen or culture
Whole genome sequencing	Entire genome	Heterogeneous SNPs	Specimen or culture
Spoligotyping	Direct repeat spacers (CRISPR)	Differences in spacer sequences based on hybridization	Specimen or culture
Phage typing[e]	Lysis of colonies by mycobacteriophages	Susceptibility to select mycobacteriophages	Culture
Pulse field gel electrophoresis[e]	Enzymatic digestion of chromosomal DNA	Differences in restriction enzyme band patterns	Culture
pTBN12 digestion[e]	Species-specific plasmid fingerprinting	Differences in restriction enzyme band patterns	Culture
hsp65 PCR-RFLP[e]	*hsp65* gene	Differences in restriction enzyme band patterns	Culture
Random amplified polymorphic DNA[e]	Species-specific chromosomal DNA	Differences in PCR band patterns	Culture
Other RFLP[e]	Species-specific DNA sequences	Differences in restriction enzyme band hybridization patterns	Specimen or culture

[a]*References for methods are present in the main text and* ***Appendix Table A2****.*

[b]*The component or property of the bacterial cell analyzed by the respective methods.*
[c]*The result obtained for interpretation using the respective methods.*
[d]*The type of material that can be used for the assay. Note that all methods except for phage typing use extracted DNA for analysis.*
[e]*Methods listed come under the "other" category in* ***Table 3.2***.

Table 3.4: Thresholds for distinguishing MSIs from microevolution events as per reviewed studies.

Year[a]	Method[b]	Microevolution[c]	MSI[c]	References
2004	*IS6110* RFLP	1-3 different bands	>3 different bands	467
	MIRU-VNTR	Double allele at 1 locus	Double alleles at ≥2 loci	
2011	MIRU-VNTR	Double allele at 1 locus	Double alleles at ≥2 loci	511
2013	WGS	<80 heterogeneous SNPs	>80 heterogeneous SNPs	313
2014	*IS6110* RFLP	1-2 different bands	>2 different bands	586
2015	WGS	≤140 heterogeneous SNPs	>140 heterogeneous SNPs	545
2019	MIRU-VNTR	Double allele at 1 locus	Double alleles at ≥2 loci	587
2019	WGS	0-5 heterogeneous SNPs	757-833 heterogeneous SNPs	544

[a]*Year of publication in chronological order.*
[b]*Details regarding specific methods are included in the main text and* ***Appendix Table A2.***
[c]*Thresholds or criteria used by the respective methods to differentiate between the two forms of infection.*

CHAPTER IV

Genomic epidemiology of *Mycobacterium avium* subsp. *paratuberculosis* isolates from Canadian dairy herds provides evidence for multiple infection events.

4.1 Abstract

Mycobacterium avium subsp. *paratuberculosis* (MAP) is the pathogen responsible for paratuberculosis or Johne's Disease (JD) in ruminants, which results in substantial economic losses worldwide. MAP transmission primarily occurs through the fecal-oral route, and the introduction of a MAP-infected animal into a herd is an important transmission route. The current study characterized MAP isolates from 67 cows identified in 20 herds from the provinces of Quebec and Ontario, Canada. Whole genome sequencing (WGS) was performed and an average genome coverage (relative to K-10) of ~14.9 fold was achieved. The total number of SNPs present in each isolate varied from 51 to 132 and differed significantly between herds. Isolates with the highest genetic variability were generally present in herds from Quebec. The isolates were broadly separated into two main clades and this distinction was not influenced by the province from which they originated. Analysis of 8 MIRU-VNTR loci and 11 SSR loci was performed on the 67 isolates from the 20 dairy herds and publicly available references, notably major genetic lineages and six isolates from the province of Newfoundland and Labrador. All 67 field isolates were phylogenetically classified as Type II (C-type). MIRU-VNTR typing showed that pattern INMV 2 (76.1%) was the most prominent pattern identified among four distinct patterns. ML-SSR typing identified 49 distinct INMV SSR patterns. The discriminatory index of the ML-SSR typing was 0.9846, which was much higher than MIRU-VNTR typing (0.3740). Although ML-SSR analysis

provides good discriminatory power, the resolution was not informative enough to determine inter-herd transmission. In select cases, SNP-based analysis was the only approach able to document disease transmission between herds, further validated by animal movement data. The presence of SNPs in several virulence genes, notably for *PE/PPE*, *mce* and *mmpL*, is expected to explain differential antigenic or pathogenetic host responses. SNP-based studies will provide insight into how MAP genetic variation may impact host-pathogen interactions. This study highlights the informative power of WGS which is now recommended for epidemiological studies and to document MGIs.

4.2 Introduction

Mycobacterium avium subsp. *paratuberculosis* (MAP) is the pathogen responsible for paratuberculosis or Johne's Disease (JD) in ruminants worldwide. Clinical expression of JD results in symptoms such as chronic diarrhea, progressive weight loss, and decreased milk production [172]. In addition, this infection has raised concerns, in part due to the possible link between MAP and Crohn's Disease (CD) in humans, leading to discussions related to the impact of MAP on chronic diseases and food safety [602–605].

MAP transmission primarily occurs through the fecal-oral route, which includes contaminated teats and milk or colostrum of infected animals allowing for direct transmission from cow to calf [167,169]. Additional sources of MAP infection include environmental contamination [170,228,605] and the introduction of infected animals into a herd [290,606]. Once transmitted, JD typically progressed into four consecutive phases [172], with reliable detection of infection only occurring with the development of clinical symptoms after a period of incubation.

On-farm biosecurity measures aim to limit the spread of infection [268]. However, the introduction of MAP into a herd of new individuals presents a risk of transmission. Due to the slow-growing, persisting nature of this disease, it is difficult to establish a clear source of infection. Molecular characterization is proving to be a powerful infection-tracking strategy and may help formulate legislation that would further aid in limiting the spread of MAP. MAP can be broadly phylogenetically classified into sheep-strains (S-type/Type I and III) and cattle-type (C-type/Type II) strains, where the latter also includes bison-type (B-type) isolates based on differences in the *IS1311* insertion sequence [274–276]. The type I, II and III classification is preferred, as it has been noted that MAP isolates from different lineages are not specific to a particular host [53,271,283–285]. Examination of the MAP genome has also allowed for the development of a variety of strain typing methods based on variable repeating elements, which may vary between strains. For example, eight MAP genomic loci are used in the MIRU-VNTR method for strain typing [287]. In comparison, analysis using the ML-SSR method consists of the examination of homopolymeric (monomeric or multimeric) repeats, 11 of which have been used for typing different MAP isolates [286]. Recently, SNP analysis-based methods have allowed for the identification of unique strains within populations at much higher resolutions than previously possible [290]. MAP is a slow-growing organism (*in-vitro* doubling time of 22-26 hours), and the mutation rate is likely to be low. Different studies have reported substitution rates ranging from 0.1-0.5 SNPs/genome/year [53,104,289,290,313].

4.3 Objectives

The purpose of this study was to examine genetic variations in MAP isolated from subclinical cows shedding the bacterium in their feces. Genome sequencing was conducted with 67 MAP isolates from 20 herds throughout the provinces of QC and ON. These provinces are the two largest dairy producers in Canada, making farms within these provinces an optimal choice for conducting a molecular epidemiology study. The use of WGS allowed for the detection of many distinct strains with significant genetic variation within a single herd, suggesting that cattle may have become infected with MAP through independent infection events, instead of originating from a single parent strain that infected a single animal at that farm. Such events have implications for infection source tracking and farm biosecurity. In addition, this study was performed in part as a proof of concept for the high-resolution mapping of MAP isolates within animals across multiple herds. Optimization of the methodology within this study allowed for validation of methods for use in examining relations of MAP at both the intra-herd level in this Chapter and the intra-cow level described in Chapter V of this thesis.

4.4 Results and Discussion

4.4.1 Animal Phenotypes and MAP Isolates

Serum ELISA measures the presence of MAP-specific antibodies, while fecal PCR is a direct measurement of the presence of MAP in the feces. While culture-based methods have the advantage of detecting live MAP, the use of quantitative PCR allows for the evaluation of bacterial shedding levels. Prior studies have compared different DNA extraction systems coupled with qPCR and evaluated the detection limit of MAP

present in feces [178]. The most efficient protocol was used to quantify fecal MAP excretion in 555 MAP-infected cows from the 3,452 cows from 22 herds that were initially tested. Samples from most animals were collected on a semi-annual basis. The specificity of fecal MSP detection is 100%, as the *ISMAP02* target is MAP-specific [125]. True MAP infections were also confirmed by serum ELISA, where specificity reached 97.1-98.6% [607,608] and fecal PCR [178]. However, because of the potential passive fecal shedding of MAP, the phenomenon of MAP being ingested from contaminated feed and detected in the feces without causing disease [609], cultures were performed on some suspected samples to confirm MAP infections. Cows with low fecal excretion of MAP with a negative serum ELISA test were confirmed to be infected with MAP using culture **(Figures 4.1 C** and **4.1 F)**. The detection limit of qPCR is equivalent to 100-500 colony-forming units (cfu)/g of feces, which corresponds to quantitative PCR values reported by the cycle threshold (Ct), of $37\text{-}38^{-\Delta\Delta CT}$ [178]. Similarly, high-shedding cows are known to excrete levels of MAP approaching 1 million cfu/g of feces, which would correspond to Ct values <27 according to the extrapolation of the standard curves **(Appendix Figure A1)** that are also similar to previous observations [178].

To obtain MAP isolates from each herd, priority was given to the cows (up to six per herd) with the highest shedding levels per herd, as detected by qPCR. Examination of fecal qPCR data of these cattle, with multiple samples recorded over a period of 3-5 years in most cases, assigned the animals to three MAP shedding categories: low, moderate, and high. Some low-shedding cows were selected for the study as they were the only MAP-shedding cows available from these farms during the longitudinal study. MAP isolates

were collected from 180 cows from 20 herds among the 376 cows analyzed by fecal culture. When analyzing feces from these cows, isolates were unable to be retrieved from two herds. Non-MAP microorganisms were able to resist the decontamination process, showing complete overgrowth of the medium with both the original culture and subsequent repeat cultures in these cases [610]. The success of the culture depends on the ability to control the outgrowth of other microorganisms which may vary depending on the diet of the cattle and geographical regions [611,612]. In addition, contamination can be grouped in samples from certain farms, which was noted in samples from the two herds where contamination was observed. Cultures were performed using aliquots of frozen MAP culture preserved at -80°C, corresponding to the highest MAP shedding period **(Figure 4.1)**. Storage of samples at -80°C and the absence of freeze and thaw cycles kept samples in optimal condition [613], allowing for the collection of MAP isolates from 4-5 years of historical samples, including from low-shedding cows **(Figures 4.1 E** and **4.1 F)**. Of the 67 cattle examined within this study **(Table 4.1)**, 42 were classified as high shedders (Ct<27), 20 were classified as moderate shedders (27<Ct<33), and five cows were classified as low shedders (Ct>33). Select examples of the profiles of each category of shedder are shown in **Figure 4.1**.

4.4.2 Phylogenetic SNP-based analysis

Of the 67 assemblies that resulted from the bioinformatic pipeline, an average genome size of ~4.78 Mb, with average GC content of 69.33% and average read coverage depth of ~14.9X (relative to K-10) was achieved **(Appendix Figure A2)**. Verification of the assemblies using CheckM confirmed that all assemblies produced were of high

quality, with average completeness of ~99.28% (classified as "near" completion) and average contamination of ~1.13% (classified as "low" contamination) [361] **(Appendix Table A4)**.

While type II isolates are typically seen in cattle, and type I/III isolates are typically found within sheep, previous studies have noted instances where these types were found in additional host organisms [271,283–285]. While it may be expected that the isolates within this study belong to the type II category due to their cattle-based origin, it is not guaranteed. To verify that the isolates examined in this study were type II strains, genetic variant analysis based on core SNP phylogenies for each isolate was performed using Snippy against the following MAP strains: Telford (type I), K-10 (type II), S397 (type III), and MAPK JB-16/15 (type B) **(Appendix Figure A2)**. All 67 field isolates from the provinces of QC and ON were shown to be closely related to K-10 **(Figure 4.2)**, with a range of 51-132 SNPs present across these 67 **(Appendix Table A5)**. The genetic variant analysis performed on these 67 isolates based on type I, III and B strains showed much higher SNP ranges of 3,333-3,429, 3,297-3,400, and 564-653, respectively **(Appendix Tables A6**, **A7** and **A8**, respectively**)**. As expected, the lower number of variants in the K-10-based genetic variant comparison indicates that these isolates are type II strains. This claim was further validated using methods established in previously published works to examine the assemblies of each isolate and verify characteristics which match type II strains. Isolates were categorized as C-type strains through the confirmation of the C-type exclusive large sequence polymorphism (LSP) $LSP^{A}20$ [272,273].

None of the isolates contained the B-type specific "TG" deletion present in *IS1311* locus 2, confirming all isolates examined in this study as being type II strains [53,614].

The total number of SNPs found varied from 51-132 SNPs **(Appendix Table A5** and significantly varied among herds **(Figure 4.3)**. Herds with isolates with the highest genetic variability were generally in the province of QC except for herd ON-3. Herd prevalence of MAP infection did not correlate (r_s=-0.19) with the genetic variability of the isolates found in herds **(Figure 4.3)**. The number of unique variants present within each assembly was also recorded **(Appendix Table A9)**. Of the 67 field isolates examined in this study, 13 of them contain no unique SNPs. The highest number of unique SNPs was 31 in strain 1452 (herd ON-2). When examining the number of unique SNPs within individual herds, all herds with multiple isolates were shown to have several isolates with their own unique MAP variant patterns.

As shown in **Figure 4.2**, the isolates are broadly separated into two main clades. This distinction is not influenced by the province from which they originate. In most cases, isolates derived from the same herds clustered together, except for herds ON-1, ON-2, ON-4, ON-6, ON-7, and ON-8. In these herds, as shown in **Figure 4.2**, the genetic diversity (standard deviation, SNP ranges) was much higher than many of the other herds examined. Interestingly, these herds were the largest of the dairy farms examined. Three of these herds (ON-6, ON-7, and ON-8) contained isolates found in both clades present. The presence of multiple distantly related strains present within individual farms suggests separate infection events with distinct MAP isolates.

In some cases, the movements of animals between herds were known and could be related to results found within the core SNP phylogeny **(Figure 4.2)**. Herds QC-2 and QC-3 are located 5 km from each other and have been trading animals for a very long time (>20 years). With recorded animal movements between the two long-standing dairy farms, it is not possible to trace the pattern of transmission between these two farms without having access to historical samples of early movements. Due to their frequent trade for decades, it is unsurprising that the isolates from these two independent commercial dairy farms cluster closely together within the core SNP phylogeny. Herds QC-5 and QC-11 cluster closely together as well, which can be partially explained by a recorded animal movement. The first MAP-positive cows in herd QC-5 were detected in 2014. A 4-year-old JD-positive cow, which isolate 392 was collected from, was purchased from herd QC-11 at the age of 3 years old and started to excrete MAP the following year. As MAP infection generally occurs in young (<2 years old) livestock [170], it suggests that this cow was already infected with MAP when purchased. The introduction of this animal to herd QC-5 also coincided with the period of births of animals that were expected to become positive in 2016–2017, most notably including the cows which isolates 885 and 1071 were collected from. This cow tracking data, along with the results observed in the core SNP tree seemingly confirm a direct ancestral transmission of MAP from herd QC-11 to QC-5.

Another potential transmission event can be found in herd ON-7. This herd contained three isolates with two distinct genetic profiles, one within isolate 631 and another within isolates 638 and 648. The animal which isolate 631 was obtained from was

born within the herd in December of 2010, while the animal which isolate 638 was extracted from was born on a separate farm in October of 2010. This second cow (isolate 638) was introduced to herd ON-7 in 2012 (at the age of 2 years, 2 months). The cow which isolate 648 was obtained from was born in 2013, leaving an age difference of ~2.5 years between this cow and the cow isolate 638 was obtained from. This timeframe would allow cow 638 to transmit its MAP strain to herd mates, including cow 648, explaining the clustering observed in **Figure 4.2**.

A third instance of potential transmission is observed in herd ON-6. This herd shows two distinct profile types, one shared by isolates 594, 1007, 1023, and 1512, and another shared by isolates 1495 and 586. The cow isolate 1512 was derived from is the oldest of these cows, which was born in December of 2008 on a separate farm from ON-6. This cow was transferred to herd ON-6 at the age of 2 years and 4 months, while cow 594, cow 1007 and cow 1023 were all born >2 years later in 2010. Cow 1512 was 5–6 years old during their birth period, allowing these three cows to become infected by cow 1512, as suggested by their similar genetic profile. Cows that resulted in isolates 586 and 1495 were born within herd ON-6, in 2010 and 2012 respectively. Cow 586 was confirmed to be shedding MAP at 5 years and 6 months old, while confirmed MAP shedding for cow 1495 was found at 3 years and 8 months old. The difference in the variant profile presented by isolates 586 and 1495 compared to the other four isolated derived from this herd, alongside the fact that both cows were born and raised within herd ON-6, suggests that a secondary infection event was present in this herd for some time.

4.4.3 Genetics and virulence factors

As previously shown in **Figure 4.1**, the shedding of MAP within host feces can vary significantly, with some hosts shedding high amounts of MAP **(Figure 4.1 A** and **4.1 B)** while other hosts shed MAP at a very low level **(Figures 4.1 E** and **4.1 F)**. The exact reason for this variation in shedding difference, whether it be due to the host, the bacteria, or an interaction between the two, is currently unknown [615]. Genetic variations in the virulence factors could impact the mechanisms allowing MAP to infect the host and thwart the mechanisms of the macrophage where it nests to ensure its survival and multiplication. Since the information available for MAP is limited, details regarding the mode of action of MAP virulence factors are imputed from functions observed in related mycobacteria, such as *M. tuberculosis*. In addition to examining the phylogeny of SNPs and transmission patterns in herds, the presence of many SNPs in several virulence gene families may help improve current knowledge of MAP virulence. The virulence factors explored in this study, namely the proline-glutamate/proline-proline-glutamate motif (*PE/PPE*) genes, mycobacterial membrane protein Large (*mmpL*), and the mammalian cell entry (*mce*) operons, were chosen based on studies on the pathogenicity of *M. tuberculosis* and *M. avium* [242,616]. The presence of genetic polymorphisms in *PE/PPE*, *mmpL*, and *mce* genes from each of our field isolates and noted variations present within these genes **(Tables 4.2, 4.3** and **4.4)**.

PE and PPE protein families are well-conserved proteins unique to mycobacteria and are suspected to have roles in the pathogenicity of the organism [617,618]. These proteins are typically found at the cell surface, with conserved N-terminal and C-terminal domains

that broadly categorize these proteins into subfamilies [619]. Some PE/PPE proteins have been shown to elicit B cell responses [617] and affect macrophage or dendritic cell function [620,621]. With some PE/PPE proteins expressed on the cell surface, one would expect to have a differential antigenic response or pathogenic variation associated with the different genetic patterns of these genes. The WGS analysis revealed several missense, frameshift, and synonymous mutations, along with a disruptive in-frame deletion in several *PE/PPE* genes **(Table 4.2)**. Genetic variation was found in only one *PE* gene, found at locus MAC_P9 (Map4144 protein). In several isolates, the deletions (missense mutations) identified would truncate the Map4144 protein. Interestingly, virulence and antigenicity of MAP during macrophage infection are dominated by the up-regulation of *Map4144* [622]. Whether the missense mutation observed affects the function of Map4144 is worth future investigation. Genetic variants were detected in 12 *PPE* genes, notable in loci MAC_PPE1, MAC_PPE7, MAC_PPE9, MAC_PPE10, MAC_PPE14, MAC_PPE16, MAC_PPE20, MAC_PPE30, MAC_PPE33, MAC_PPE33, and MAC_PPE41 **(Table 4.2)**. All of these genes are conserved in other MAC bacteria, though the exact function of each protein is uncertain. It has also been noted that seven of these proteins have no orthologue in *M. tuberculosis* [623], with the remaining six loci corresponding to Map0123 (MAC_PPE1), Map2575c (MAC_PPE10), Map1518 (MAC_PPE20), Map3185 (MAC_PPE33), MAP3725 (MAC_PPE41) and Map1734 (uncharacterized MAP locus) proteins, have orthologues in *M. tuberculosis*, corresponding to PPE20, PPE18, PPE32, PPE51, PPE3, and PPE37, respectively. Given the importance of the PE/PPE family for virulence, an importance which also extends to any mutations that could affect the host

antigenic response, further investigation into the PE/PPE genes with mutations identified in this study is recommended for future studies.

Factors that enable the pathogen to achieve infection and allow persistence are also considered to be virulence factors. Genetic variations in mammalian cell entry (mce) operons, implicated in the invasion of mycobacteria into host cells, are expected to influence MAP virulence as well. MAP contains eight *mce* operons instead of the four present in *M. tuberculosis*, with two copies of the *mce5* and *mce7* operons [624,625]. The structure of *mce* genes in each operon is composed of eight genes (*mceA-mceF* and *yrbEA-B*). Several missense mutations were identified in *mce1A*, *mce1F*, *mce5-2E*, *mce5-2F*, *mce7-2F* and *yrbE1B* and *yrbE2B* genes though their impact on MAP pathogenicity is unknown at this time **(Table 4.3)**.

The cell envelope is a critical aspect of mycobacterial physiology and virulence. Mycobacterial membrane protein Large (mmpL) proteins are permeases which are an integral part of the membrane [626]. These membrane transporters indirectly contribute to virulence in *M. tuberculosis* via the transport of substrates that directly influence bacterial survival. While information regarding the function of these mmpL proteins is more detailed for *M. tuberculosis* [626], their role in MAP is poorly understood [616]. Future investigation into the role of mmpL proteins in MAP, alongside how the mutations listed in **Table 4.4** may impact protein function, is recommended for future studies,

4.4.4 *In-silico* analysis of repetitive elements

VNTRs such as those described in the MIRU or ML-SSR methods [286,287] are a significant source of genetic polymorphisms in mycobacteria, including MAP. Due to the

low mutation rate of this slow-growing organism, repetitive DNA elements become one of the main driving forces of genome evolution within this species, Multi-locus variant analysis (MLVA) methods provide a discriminatory power and, more importantly, the portability of the results (MLVA codes) facilitates their exchange between laboratories. *In-silico* analysis of eight MIRU-VNTR loci, namely MIRUs 292 and X3, and VNTRs 25, 47, 3, 7, 10 and 32 [287], and 11 SSR loci [286] was performed on both the 67 isolates from 20 dairy herds, as well as ten strains downloaded from the NCBI genome database: Telford (type I), K-10 (type II), S397 (type III), MAPK JB16/15 (B-type), and six isolates previously found on the province of Newfoundland and Labrador (NL), Canada (NL-89C, NL-93B, NL-95A, NL-95B, NL-95E, NL-96E) [289,315]. The repeat size at each locus was recorded manually for each isolate, along with the herd ID, province ID, and MAP shedding level of the 67 cows each isolate was sampled from **(Table 2.2)**.

Among the 67 isolates examined with SSR markers, loci 1 and 2 showed the highest discriminatory index (DI) values at 0.8508 and 0.8005, respectively **(Table 2.2)**. For SSR locus 1, a total of 15 distinct repeat patterns were identified, while SSR locus 2 had a total of eight distinct repeat patterns **(Table 2.2)**. Other SSR loci had only one repeat pattern, including SSRs 3, 4, 5, 10, and 11, thus excluding their discriminatory capability. Apart from the high discriminatory power of loci 1 and 2, SSR loci 7 (DI of 0.4242) displayed three different patterns, followed by loci 6, 8, and 9 with a much lower DI (0.0299 each), with only two patterns. Among the MIRU-VNTR loci examined, locus 10 was the most discriminatory locus (DI of 0.2985), followed by locus 292 (DI of 0.0882) and locus 7 (0.0298), all with only two different repeat patterns present within the

67 field isolates examined. The remaining MIRU-VNTR loci were not informative. All MIRU-VNTR *in-silico* results matched those obtained by PCR (data not shown). As shown in **Figure 4.4**, ML-SSR markers were more informative than the MIRU-VNTR markers. The DI values reported for the MIRU-VNTR markers were globally lower than those found in the SSR markers, as previously shown in other studies [627,628].

Examination of the MIRU-VNTR and SSR loci was also performed using the web application "MAC-INMV-SSR" database, a database which contains the MIRU-VNTR and SSR profiles of several known strains of MAP [373]. Using this database to examine the 67 type II isolates from the provinces of QC and ON, four different INMV-MIRU patterns were identified **(Table 2.2)**. The predominant type of field isolates in these provinces was INMV 2 (76.1%), followed by INMV 3 (17.9%), INMV 13 (4.5%) and INMV 8 (1.5%). The total DI using the INMV database patterns is 0.3740.

Analysis of the INMV-SSR pattern was obtained for 67 isolates from QC and ON. Some strains (n=8) had an ambiguous number of repeats in their assembly (for the field isolates assembled) or a lack of coverage over the entire repeat (for the reference sequences downloaded from NCBI). Manual examination of the repeats allowed for the confirmation of the SSR patterns of six of the field isolate assemblies, leaving the remaining two field isolates, along with two NCBI sequences, listed as having an unknown pattern. Of the remaining 74 isolates, 15 recognized patterns were detected, along with 16 novel patterns which were not present within the MAC-INMV-SSR database. An important point of consideration is that the MAC-INMV database has a slight limitation when examining SSR loci. If the repeat value is greater than 11, typically

shown within the longer mononucleotide repeats in SSR loci 1 and 2, then the repeat is called as "+". This results in a slight underestimation of the true diversity of the population. Notwithstanding, the stability of the SSR locus 2 has previously been shown to be affected by *in-vivo* passage [629], which may not be suitable for epidemiological study. Four of the patterns within the database (SSR7, SSR8, SSR10, and SSR93), which called long repeats as "+", were found to have several repeat "sub-patterns" within each called pattern (two in SSR7, two in SSR8, six in SSR10, and five in SSR93), showing further diversity within listed repeat patterns. Among the 16 novel patterns, six were found to have additional patterns that would not be recognized due to their high repeat values. When taking these additional patterns into account, the DI value increases from 0.9457 to 0.9846, which is reported in **Tables 2.2** and **4.4**.

The INMV-MIRU-VNTR DI value of the 67 isolates was evaluated at 0.3740 while the INMV-SSR was 0.9846. The capacity of INMV-SSR to differentiate MAP isolates exceeds that established using the alternative method previously reported (DI if 0.795), which used a selection of novel tandem repeats [616], making the web application "MAC-INMV-SSR" database the best reference for reporting typed MAP.

Comparison of both ML-SSR and MIRU-VNTR patterns, as shown in both **Table 4.5** and **Figure 4.4**, allows for a clear visualization of how these patterns are spread throughout the isolates examined. Among the 18 herds with multiple isolates, MIRU-VNTR was only able to identify multiple patterns in herds ON-2 and ON-6, each confirmed as having two patterns. All other herds contained a single INMV type. Using SSR loci, multiple patterns were confirmed in 17 of these 18 herds, with a minimum of

two confirmed profiles (seven herds) and a maximum of six confirmed profiles (ON-6). A single pattern (SSR13) was confirmed in the herd QC-9 (isolate 1427). While the isolates were shown to be closely related by core-SNP analysis using ~107 SNPs, each isolate was confirmed to have a unique pattern of variants, with no clonal isolates identified. The differentiation provided by both core-SNP analysis and ML-SSR typing supports previous work which has noted that repetitive elements, especially MIRU-VNTR, are often subject to homoplasy [53,630]. In these cases, while the MIRU-VNTR pattern may be the same, further examination would reveal that the isolations are genetically distinct from one another and are unrelated. Homoplasy was obvious within the isolates examined in our study, as the broad classification of unrelated isolates as INMV 2 indicates. Homoplasy within INMV 2, itself a commonly identified MIRU-VNTR type in both Europe and Canada [284,290,630], was also noted as being an example of homoplasy in prior works [53,630].

4.5 Conclusion

In summary, MAP isolates from 67 cattle in 20 herds within the provinces of QC and ON were genetically characterized using a selection of typing methods, including WGS analysis. The assemblies produced using our prospective pipeline were verified as being of high quality without contamination, which avoids the generation of misleading hypotheses. All isolates were type II, the cattle type. They were subtyped using both traditional MLVA typing methods such as MIRU-VNTR and ML-SSR. Examination of the SSR and MIRU-VNTR-based methods showed that SSR had significantly more discriminatory power than MIRU-VNTR, as it was able to identify more unique patterns both across the entire dataset and within each of the individual herds. However, the

limited discriminatory capacity of the multi-locus typing methods does not allow tracking of inter-herd MAP transmission. The core SNP-based analysis was the only approach leading to the assignment of individual signatures to each isolate and allowed for the documentation of disease transmission across herds, supported and confirmed by the animal movement data. While the structure of the 67 isolates within the phylogenetic tree was not correlated with the phenotypes recorded in the animals, the presence of genetic variations in several virulence genes, notable for *PE/PPE*, *mce*, and *mmpL* genes, could explain differential antigenic or pathogenetic responses, though further analysis of these SNPs is required to verify this hypothesis. Additional isolates are being studied and the analysis of a larger population should make it possible to study the impact of the genetic pattern of MAP virulence genes and their impact on the host in a context and controlled infection.

4.6 Figures and Tables

4.6.1 Figures

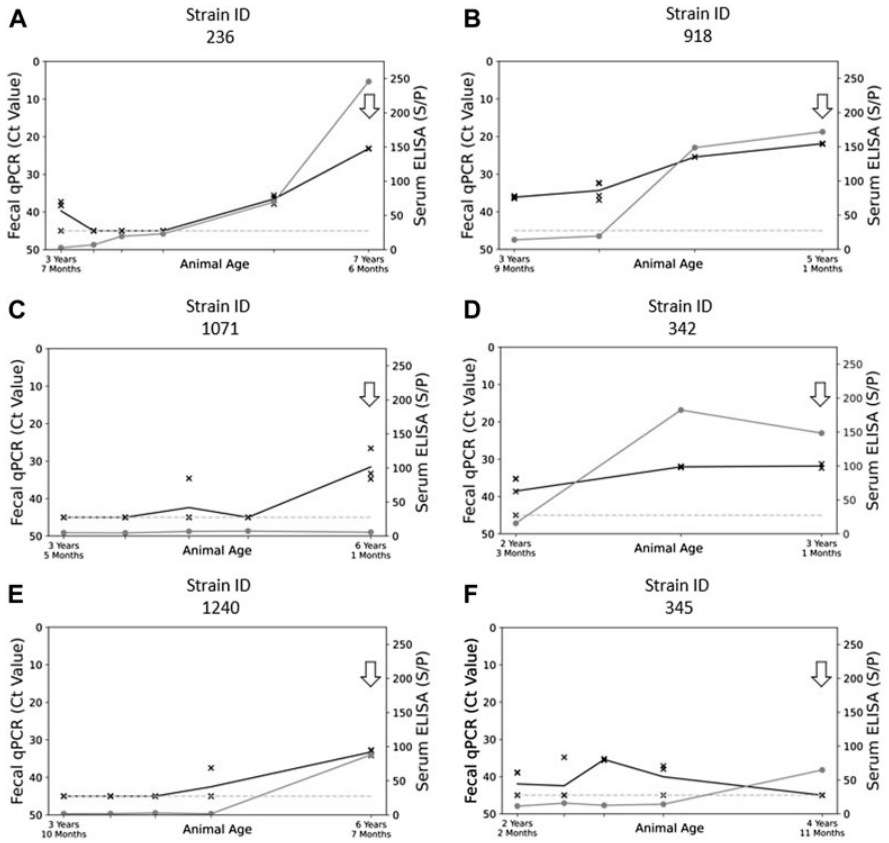


Figure 4.1: Graphical plots showing the relationship of fecal shedding of MAP (qPCR Ct value) and blood ELISA (Sample/Positive Ratio) to animal age. Panels **(A)** and **(B)** reflect the profile of phenotypes collected in "high shedding" animals (Ct <27), panels **(C)** and **(D)** reflect the trends present in "moderately shedding" animals (27<Ct<33), and panels **(E)** and **(F)** reflect the trends present in "low shedding" animals (Ct>33). The arrows present on each graph represent the timepoint in each animal that fecal samples were selected for the culture of MAP. The grey line with points represented by "O" represents the S/p-value of the serum ELISA at that point in time. The black line with points represented by "X" represents the Ct results (mean value) obtained from qPCR at that point in time. The dashed line along the bottom represents the minimum ELISA threshold for isolates to be suspected of being MAP-positive.

Figure 4.2: Core SNP phylogeny of Type II MAP strains. Phylogenetic tip labels are coloured by herd. Tip labels in grey represent type II strains used as reference sequences downloaded from the NCBI database. The isolates are divided into two major clades, represented by the main axis which is divided into two major branches. Arrows leading from one isolate to another are indicative of a confirmed transmission event as described in Section 4.4.2. Arrows with arrowheads at both ends are present at herds QC-2 and QC-3 indicating that while transfer between herds has likely occurred, the exact transmission pattern of MAP between hosts is unknown.

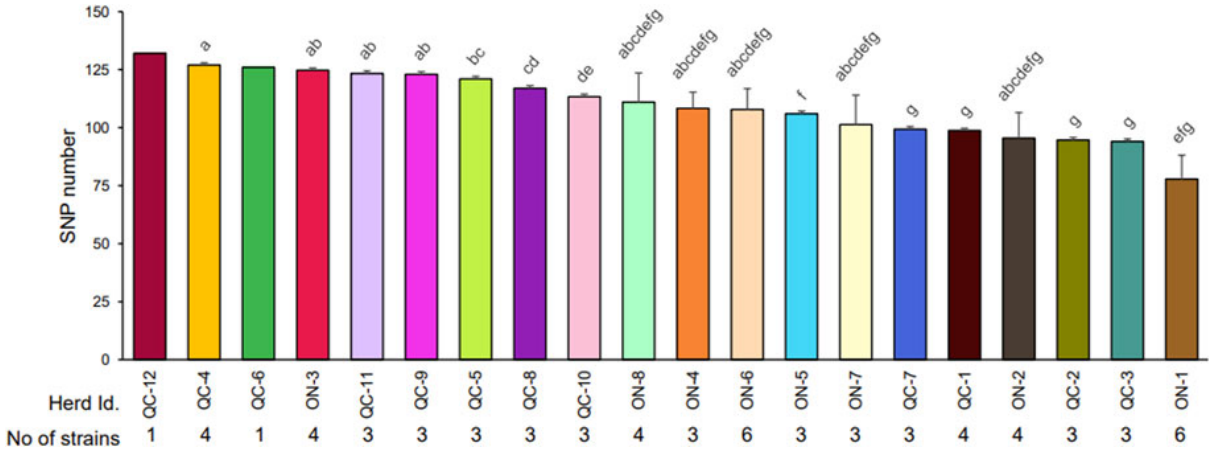


Figure 4.3: Average number of SNPs detected in the strains isolated within each herd. The number of SNPs in each strain was recorded **(Appendix Table A5)**. The average number of SNPs detected in strains isolated in each herd with different letters differ significantly ($p<0.05$) after the application of a Tukey correction. The number of strains analyzed by WGS is reported for each herd. Each strain is derived from a single isolated colony from a different cow. Statistical analysis of SNP counting was performed by Dr. N. Bissonnette.

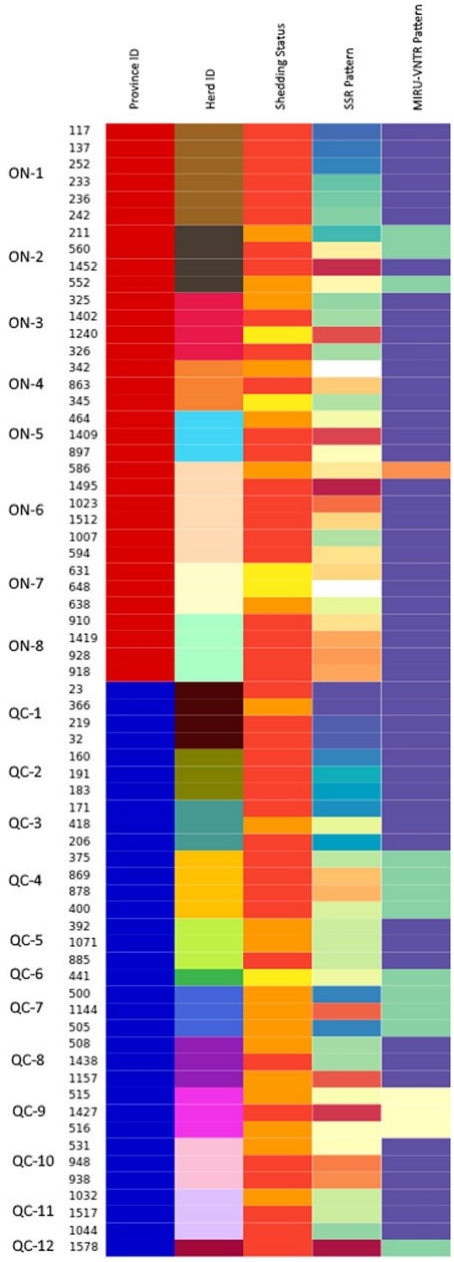


Figure 4.4: Visualization of the distribution of shedding status, SSR type and MIRU-VNTR type across different herds in ON and QC. Isolates are grouped by herd, with colours in each column reflecting a different value within their respective category. In the province column, isolates from QC are labeled in blue, while isolates from ON are labeled in red. Herd colours are the same as those in **Figures 4.2** and **4.3**, and are labelled long the side of the figure. The shedding status column identifies isolates from animals that were high shedders (red), moderate shedders (orange) and low shedders (yellow). Isolates with no shedding data were coloured grey. Both the SSSR type and MIRU-VNTR type columns use colours to represent patterns, with unknown patterns coloured white.

4.6.2 Tables

Table 4.1: Herd size, prevalence of infection per-herd and samples examined per-herd for 14 herds examined.

Herd ID	QC-1	QC-2	QC-3	QC-4	QC-5	QC-6	QC-7	QC-8	QC-9	QC-10	QC-11	QC-12	ON-1	ON-2	ON-3	ON-4	ON-5	ON-6	ON-7	ON-8
Herd Size[a]	157	71	90	91	45	47	82	265	121	156	195	104	423	231	162	155	122	378	208	155
Prevalence (%)[b]	7.0	29.6	10.0	2.2	15.5	6.4	14.7	11.7	18.2	19.2	4.1	16.4	11.1	11.7	11.1	8.4	4.9	32.5	10.1	24.6
MAP-infected cows	11	21	9	2	7	3	12	31	22	30	8	17	47	27	18	13	6	123	21	38
Number of strains[c]	4	3	3	4	3	1	3	3	3	3	3	1	6	4	4	3	3	6	3	4

[a]All cows in this study were older than 24 months. Herd size reflects the number of cows over 24 months old when blood and feces were collected at the first visit to the dairy farm.
[b]Prevalence (%) represents the true prevalence of MAP-infected cows recorded during the longitudinal study of 3-5 years.
[c]The number of strains analyzed by whole genome sequencing (WGS) is reported for each herd. Each strain is derived from a single isolated colony from the selected cow.

Table 4.2: Mutations identified in PE/PPE proteins in reference to the MAP K-10 genome.

PE/PPE Name[a]	MAP gene/protein (K-10)	TB Orthologue	Location of Variant[b]	Variant Type[c]	Nucleotide Alteration	Protein Alteration	Variant Effect	Herd ID[d]
MAC_PPE1	Map0123	PPE20	Between 131778 - 131779	INS	+CGG	Ala duplication	Disruptive Inframe Insertion	QC-2 (191), QC-3 (206), ON-2 (560)
MAC_PPE30	Map1144c	N/A	1198312	SNP	C -> G	Gln -> His	Missense mutation	QC-5, QC-8 (1157), QC-11
MAC_PPE20	MapP1518	PPE32	1668384 - 1668404	DEL	- CGCCGCC TATGAGAC GGCCTT	Shortening of Protein	Conservative Inframe Deletion	ON-3 (325)
MAC_PPE16	Map1676	N/A	1831523	SNP	G -> A	Gly -> Glu	Missense mutation	ON-6 (594, 1023, 1512), ON-7 (631), ON-8 (910, 918, 1419)
UNK	Map1734	PPE37	Between 1896185 - 1896186	INS	+C	Shortening of protein	Frameshift mutation	ON-7 (638, 648)

MAC_PPE14	Map1813c	N/A	1990165 - 1990219	DEL	- GCCGCGG CGGCGCC GACGGGC CCGCGCC TTCTTGCT GCCCGCG GCCGCCG CCGCG	Shortening of protein	Frameshift mutation	ON-6 (594, 1512)
MAC_PPE14	Map1813c	N/A	1991288	SNP	G -> A	Ala ->Val	Missense mutation	QC-5, QC-8 (1157), QC-11
MAC_PPE10	Map2575c	PPE18	2896134 - 2896145	DEL	- CGCCGCC GACCG	Shortening of Protein	Disruptive Inframe Deletion	QC-10
MAC_PPE9	Map2595	N/A	2920978	DEL	-A	Shortening of protein	Frameshift mutation	All Isolates
MAC_PPE9	Map2595	N/A	2921144	SNP	C -> T	Pro -> Leu	Missense mutation	QC-2, QC-3
MAC_PPE7	Map2601	N/A	2927781	SNP	G -> A	Gly -> Arg	Missense mutation	QC-5, QC-11

MAC_PPE33	Map3185	PPE51	3536540	SNP	C -> T	Leu -> Leu	Synonymous mutation	QC-5, QC-11
MAC_PPE36	Map3490	N/A	Between 3881083 and 3881084	INS	+G	Shortening of protein	Frameshift mutation	QC-4 (400)
MAC_PPE41	Map3725	PPE3	4146287	SNP	C -> T	Asn -> Asn	Synonymous mutation	ON-6 (594, 1023, 1512), ON-7 (631), ON-8 (910, 918, 1419)
MAC_PE9	Map4144	N/A	4622682	DEL	-T	Shortening of protein	Missense mutation, "intergenic region"	QC-1, QC-2 (183), QC-3 (171), QC-4 (375, 400, 869), QC-5, QC-6, QC-8 (1157), QC-9 (516), ON-2 (211, 1452), ON-3 (325, 326), ON-7 (638, 648)

MAC_PE9	Map4144	N/A	4622687	DEL	-T	Shortening of protein	Missense mutation, "intergenic region"	QC-1, QC-2 (183), QC-4 (375), QC-5 (392), QC-9 (515, 516), ON-3 (325, 326), ON-7 (648), QC-8 (1157)
MAC_PE9	Map4144	N/A	4622694	DEL	-A	Shortening of protein	Missense mutation, "intergenic region"	QC-1, QC-2 (183), QC-3 (171), QC-4 (375, 869), QC-5 (392), QC-6, QC-9 (515, 516), ON-3 (325), ON-7 (648), ON-8 (1419)

[a]Proteins identified as described in Mackenzie et al. (2009)[623] (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2668356/). One protein, an orthologue of M. tuberculosis protein PPE 20, was not given an annotation by this publication and is listed as unknown (UNK).

[b]Mutations are sorted according to nucleotide locus in reference to the K-10 genome sequence (Accession ID: AE016958.1)

[c]SNP = Single Nucleotide Polymorphism, INS = Insertion, DEL = Deletion

[d]Listing of the herd ID indicates that all isolates taken from that herd have the same mutation. If only certain isolates from a herd contain the mutation, they are listed in parenthesis.

Table 4.3: Mutations identified in mce proteins in reference to the MAP K-10 genome.

mce Gene[a]	MAP gene/protein (K-10)	Location of Variant (K-10)[b]	Variant Type[c]	Nucleotide Alteration	Protein Alteration	Variant Effect	Herd ID[d]
mce7-2F	Map0113	124194	SNP	C -> T	Ile -> Ile	Synonymous mutation	QC-3 (171)
mce7-2F	Map0113	124251	SNP	G -> T	Lys -> Asn	Missense mutation	QC-12
mce4F	Map0569	594004	SNP	T -> C	Leu -> Pro	Missense mutation	QC-1, QC-2, QC-3 (171), QC-7, ON-1 (117, 137, 242), ON-2 (211, 552), ON-4 (342, 863), ON-5, ON-6 (586, 1495), ON-7 (638, 648)
mce3E	Map2112c	2338751	SNP	G -> A	Pro -> Pro	Synonymous mutation	ON-5

mce5-2E	Map2193	2436439	SNP	C -> T	Pro -> Leu	Missense mutation	QC-5, QC-8, QC-11
mce5-2E	Map2193	2436501	SNP	G -> T	Ala -> Ser	Missense mutation	QC-1, QC-2 (183, 191), QC-3 (206, 418), QC-4 (375, 400, 869), QC-5 (885, 1071), QC-7, QC-8 (1157, 1438), QC-9, QC-10 (938, 948), ON-1 (117, 233, 236, 252), ON-2 (211, 552, 560), ON-3 (326, 1240), ON-4 (342), ON-5, ON-6 (594, 1495, 1512), ON-7 (631), ON-8

mce5-2F	Map2194	2437285	SNP	G -> C	Val -> Leu	Missense mutation	ON-1 (117, 137 and 242)
mce5-2F	Map2194	2437484	SNP	C -> G	Ala -> Gly	Missense mutation	All strains
Independent mce	Map3289c	3652375	SNP	G -> C	Pro -> Pro	Synonymous mutation	QC-6
Independent mce	Map3289c	3652383	SNP	T -> G	Ile -> Leu	Missense mutation	QC-6
Independent mce	Map3289c	3653470	SNP	C -> G	Thr -> Thr	Synonymous mutation	QC-6
yrbE1B	Map3603	3999143	DEL	-G	Extension of protein	Frameshift mutation	All Isolates
mce1A	Map3604	4000027	SNP	C -> G	Thr -> Thr	Synonymous mutation	QC-10 (938)
mce1A	Map3604	4001149	SNP	G -> C	Pro -> Pro	Synonymous mutation	QC-5, QC-8, QC-11
mce1F	Map3609	4006938	SNP	C -> G	His -> Gln	Missense mutation	QC-5 (885)
yrbE2B	Map4083	4553018	SNP	T -> C	Tyr -> His	Missense mutation	QC-5, QC-8, QC-11

yrbE2B	Map4083	4553276 - 4553278	DEL	-GTG	-Val, Shortening of protein	Disruptive Inframe Deletion	QC-9, ON-3 (326, 1240), ON-4 (345)
yrbE2B/mce2A	Map4083/ Map4084	4553474	DEL	-C	Extension of protein mce2A	Frameshift mutation	All isolates
mce2D	Map4087	4557531	SNP	T -> C	Val -> Ala	Missense mutation	All isolates
Independent mce	Map3289c	3653477 - 3653478	Complex	GC -> TT	Ala -> Asn	Missense mutation	QC-6
mce1A	Map3604	4000019 - 4000020	MNP	AA -> GC	Asn -> Ala	Missense mutation	QC-10 (938)

[a]*Proteins identified as described in Hemati et al. (2019)* [625] *(https://pubmed.ncbi.nlm.nih.gov/31282842/).*
[b]*Mutations are sorted according to nucleotide locus in reference to the K-10 genome sequence (Accession ID: AE016958.1)*
[c]*SNP = Single Nucleotide Polymorphism, INS = Insertion, DEL = Deletion, Complex = Combination of SNP/multi-nucleotide polymorphism*
[d]*Listing of the herd ID indicates that all isolates taken from that herd have the same mutation. If only certain isolates from a herd contain the mutation, they are listed in parenthesis.*

Table 4.4: Mutations identified in mmpL proteins in reference to the MAP K-10 genome.

MAP gene/protein (K-10)[a]	Location of Variant (K-10)[b]	Variant Type[c]	Nucleotide Alteration	Protein Alteration	Variant Effect	Herd ID (Isolate)[d]
Map1738	1899360	SNP	C -> T	Ala -> Val	Missense mutation	ON-6 (594, 1023, 1512)
Map2232	2494145	SNP	C -> T	Ala -> Ala	Synonymous mutation	QC-4, QC-6, QC-12, ON-2 (560)
Map2239	2502885	SNP	A -> G	Asp -> Gly	Missense mutation	QC-2, QC-3
Map2324c	2606590	SNP	A -> T	Leu -> Gln	Missense mutation	ON-8 (910, 918, 1419)
Map2324c	2606598	SNP	G -> A	Thr -> Thr	Synonymous mutation	QC-4, QC-6, QC-12, ON-2 (560)
Map2324c	2607271	SNP	G -> A	Ala -> Val	Missense mutation	QC-9, ON-3, ON-4 (345)

Map2324c	2607913	SNP	T -> C	Asn -> Ser	Missense mutation	ON-6 (594, 1023, 1512), ON-7 (631), ON-8 (910, 918, 1419)
Map3049c	3393600	SNP	G -> A	Pro -> Ser	Missense mutation	QC-1, ON-6 (1495)
Map3751	4187170	SNP	A -> G	Asn -> Asp	Missense mutation	QC-10
Map3751	4188170	DEL	-A	Shortening of protein, conversion of initial gene into two proteins annotated as "MMPL family transporter CDS"	Frameshift mutation	ON-1 (117, 137, 242)
Map3890	4354347	SNP	G -> A	Asp -> Asn	Missense mutation	ON-7 (631)

[a]*Proteins identified as described in Marri et al. (2006)* [631] *(https://pubmed.ncbi.nlm.nih.gov/17064286/).*
[b]*Mutations are sorted according to nucleotide locus in reference to the K-10 genome sequence (Accession ID: AE016958.1)*

[c]*SNP = Single Nucleotide Polymorphism, INS = Insertion, DEL = Deletion, Complex = Combination of SNP/multi-nucleotide polymorphism*

[d]*Listing of the herd ID indicates that all isolates taken from that herd have the same mutation. If only certain isolates from a herd contain the mutation, they are listed in parenthesis.*

Table 4.5: ML-SSR and MIRU-VNTR Patterns according to the MAC-INMV database.

Strain ID[a]	Herd ID[b]	Province	Shedding Status	INMV SSR Type[c]	INMV MIRU Type[c]
23	QC-1	QC	High	MLSSR 10a	INMV 2
32	QC-1	QC	High	MLSSR 38	INMV 2
117	ON-1	ON	High	N/P 12a	INMV 2
137	ON-1	ON	High	N/P 15	INMV 2
160	QC-2	QC	High	MLSSR 10d	INMV 2
171	QC-3	QC	High	MLSSR 10c	INMV 2
183	QC-2	QC	High	MLSSR 8a	INMV 2
191	QC-2	QC	High	MLSSR 10b	INMV 2
206	QC-3	QC	High	MLSSR 8a	INMV 2
211	ON-2	ON	Moderate	MLSSR 8b	INMV 3
219	QC-1	QC	High	MLSSR 38	INMV 2
233	ON-1	ON	High	MLSSR 93e	INMV 2
236	ON-1	ON	High	MLSSR 93b	INMV 2
242	ON-1	ON	High	N/P 12b	INMV 2

252	ON-1	ON	High	MLSSR 10d	INMV 2
325	ON-3	ON	Moderate	MLSSR 17	INMV 2
326	ON-3	ON	High	MLSSR 11	INMV 2
342	ON-4	ON	Moderate	UNK	INMV 2
345	ON-4	ON	Low	N/P 3a	INMV 2
366	QC-1	QC	Moderate	MLSSR 10a	INMV 2
375	QC-4	QC	High	N/P 14	INMV 3
392	QC-5	QC	Medium	MLSSR 13	INMV 2
400	QC-4	QC	High	N/P 10b	INMV 3
418	QC-3	QC	Moderate	MLSSR 10e	INMV 2
441	QC-6	QC	Low	N/P 11a	INMV 3
464	ON-5	ON	Moderate	MLSSR 93c	INMV 2
500	QC-7	QC	Moderate	MLSSR 10d	INMV 3
505	QC-7	QC	Moderate	MLSSR 10d	INMV 3
508	QC-8	QC	Moderate	MLSSR 11	INMV 2
515	QC-9	QC	Moderate	MLSSR 16	INMV 13

516	QC-9	QC	Moderate	N/P 6a	INMV 13
531	QC-10	QC	Moderate	MLSSR 10f	INMV 2
552	ON-2	ON	Moderate	MLSSR 7a	INMV 3
560	ON-2	ON	High	N/P 8	INMV 3
586	ON-6	ON	Moderate	MLSSR 93a	INMV 8
594	ON-6	ON	High	N/P 4a	INMV 2
631	ON-7	ON	Low	N/P 1	INMV 2
638	ON-7	ON	Moderate	MLSSR 10e	INMV 2
648	ON-7	ON	Low	UNK	INMV 2
863	ON-4	ON	High	N/P 6b	INMV 2
869	QC-4	QC	High	N/P 10a	INMV 3
878	QC-4	QC	High	N/P 13	INMV 3
885	QC-5	QC	High	MLSSR 13	INMV 2
897	ON-5	ON	High	MLSSR 10f	INMV 2
910	ON-8	ON	High	N/P 4a	INMV 2
918	ON-8	ON	High	N/P 2	INMV 2

928	ON-8	ON	High	N/P 16	INMV 2
938	QC-10	QC	High	MLSSR 93d	INMV 2
948	QC-10	QC	High	N/P 9b	INMV 2
1007	ON-6	ON	High	N/P 3a	INMV 2
1023	ON-6	ON	High	N/P 4b	INMV 2
1032	QC-11	QC	Moderate	MLSSR 13	INMV 2
1044	QC-11	QC	High	MLSSR 17	INMV 2
1071	QC-5	QC	Moderate	MLSSR 13	INMV 2
1140	QC-7	QC	Moderate	MLSSR 7b	INMV 3
1157	QC-8	QC	Moderate	N/P 3b	INMV 2
1240	ON-3	ON	Low	N/P 3c	INMV 2
1402	ON-3	ON	High	MLSSR 11	INMV 2
1409	ON-5	ON	High	N/P 9a	INMV 2
1419	ON-8	ON	High	N/P 2	INMV 2
1427	QC-9	QC	High	MLSSR 58	INMV 13
1438	QC-8	QC	High	MLSSR 11	INMV 2

1452	ON-2	ON	High	MLSSR 28	INMV 2
1495	ON-6	ON	High	N/P 5	INMV 2
1512	ON-6	ON	High	N/P 1	INMV 2
1517	QC-11	QC	High	MLSSR 13	INMV 2
1578	QC-12	QC	High	N/P 7	INMV 3
K-10	NCBI	N/A	N/A	MLSSR 10d	INMV 3
NL-89C	NCBI	NL	N/A	MLSSR 3	INMV 68
NL-93B	NCBI	NL	N/A	MLSSR 54	UNK
NL-95A	NCBI	NL	N/A	MLSSR 52	UNK
NL-95B	NCBI	NL	N/A	MLSSR 17	N/P 1
NL-95E	NCBI	NL	N/A	MLSSR 17	UNK
NL-96E	NCBI	NL	N/A	MLSSR 38	UNK
MAPK_JB16/15	NCBI	N/A	N/A	MLSSR 29	UNK
S397	NCBI	N/A	N/A	UNK	UNK
Telford	NCBI	N/A	N/A	UNK	INMV 72
Discriminatory Index				0.9846	0.3740

[a]Typing of isolates from Newfoundland Canada (NL-89C/93B/95A/95B/96E) were published (http://www.ncbi.nlm.nih.gov/pubmed/25927612).
[b]Values from the NCBI database are not included in DI calculations.
[c]N/P = New Patterns. Patterns are not recognized as a type in the INMV database. Patterns with the same number but different letters would be classified as the same type in the INMV database but have different repeat values. UNK = Unknown. These patterns can't be predicted due to a lack of certainty within specific loci and were not included in DI calculations.

CHAPTER V

Investigating *in-vivo Mycobacterium avium* subsp. *paratuberculosis* microevolution and mixed strain infections

5.1 Abstract

Mycobacterium avium subsp. *paratuberculosis* (MAP) causes Johne's Disease (JD) in ruminants, which is responsible for significant economic loss to the global dairy industry. Mixed strain infection (MSI) refers to the concurrent infection of a susceptible host with genetically distinct strains of a pathogen, whereas within-host changes in an infecting strain leading to genetically distinguishable progeny is called microevolution. The two processes can influence host-pathogen dynamics, disease progression and outcomes, but not much is known about their prevalence and impact on JD. Therefore, we obtained up to 10 MAP isolates each from 14 high-shedding animals and subjected them to whole genome sequencing. Twelve of the 14 animals examined showed evidence for the presence of MSIs and microevolution, while the genotypes of MAP isolates from the remaining two animals could be attributed solely to microevolution. All MAP isolates that were otherwise isogenic had differences in short sequence repeats (SSRs), of which SSR1 and SSR2 were the most diverse and homoplastic. Variations in SSR1 and SSR2, which are located in *ORF1* and *ORF2*, respectively, affect the genetic reading frame, leading to protein products with altered sequences and computed structures. The ORF1 gene product is predicted to be a MAP surface protein with possible roles in host immune modulation, but nothing could be inferred regarding the function of ORF2. Both genes are conserved in *Mycobacterium avium* complex members, but SSR1-based modulation of *ORF1* reading frames seems to only occur in MAP, which could have potential implications on the infectivity of this pathogen.

5.2 Introduction

Mycobacterium avium subsp. *paratuberculosis* (MAP) causes paratuberculosis or Johne's Disease (JD) in ruminants; a disease which is associated with chronic diarrhea, progressive weight loss and decreased milk production in cattle [172]. The disease is responsible for significant economic losses to the dairy industry due to reduced milk production, increased cattle mortality, and culling of infected animals, as there is currently no effective vaccine or treatment for JD [9]. A better understanding of MAP pathogenicity and transmission mechanisms is also important from a public health perspective due to the potential link between MAP and Crohn's disease in humans, and concerns have been raised regarding the impact of MAP on other chronic diseases and food safety [602–605]. MAP is transmitted via the fecal-oral route [161,162] and can survive for extended periods in the environment [225,632]. Therefore, JD control strategies must take a holistic approach to better understand MAP evolution dynamics, the diversity of strains circulating throughout a herd, and the impact of strain diversity on disease progression.

MAP is a member of the *Mycobacterium avium* complex (MAC), a group of slow-growing non-tuberculous mycobacteria, which are opportunistic pathogens of humans and animals [39,633,634]. Of the other subspecies of *M. avium*, MAP is most closely related to *M. avium* subsp. *hominissuis*, which typically infects pigs and humans [61,633–635]. MAP strains are phylogenetically classified as either sheep-type (S-type/Type I and III) or cattle-type (C-type/Type II) strains, with the "C-type" also comprising bison-type (B-type) isolates [274–276]. Classification of strains using the type I, II and III system is preferred as MAP isolates from "S-type" or "C-type" lineages are not limited to causing infections in sheep and cattle species, respectively [53,271,283–285]. Over the years, additional MAP strain typing methods have been developed, which are based on analyzing differences in the copy numbers or sequences of

variable DNA repeats. These include the eight locus MIRU-VNTR, and the eleven locus ML-SSR typing methods, both of which are commonly used in epidemiology and source tracking studies [286,287]. More recently, WGS followed by SNP analysis has allowed for strain typing at a much higher resolution than previously possible, allowing for a greater understanding of the MAP genome [53,288,290].

Mixed strain infections (MSIs) refer to polyclonal or co-infections involving strains of a single pathogenic species [468,636,637]. In some pathogens, an initially infecting strain may undergo within-host evolution to give rise to genetically distinct isolates, which is referred to as microevolution [297–299,310]. The presence of genetically distinct variants of a single pathogen species in an infected host, whether it be due to MSI or microevolution, is collectively referred to as mixed genotype infection (MGI) **(Figure 1.2)**. Such events are expected to influence the progression and outcome of an infection, along with transmission dynamics, due to differences in the physiological characteristics of isolates [296,306–309]. In some pathogens, MGIs lead to treatment complications due to the presence of multiple antibiotic resistance profiles (also referred to as heteroresistance) [300,301]. There is no approved treatment against JD [170] so the evolutionary pressure on MAP is not associated with antibiotic exposure but instead arises from prolonged infection of the host. Most cattle with JD are infected as calves at six months of age or younger and often display disease symptoms years later [163,172]. The genome evolution rate of MAP is low, with estimates ranging from 0.1 to 0.5 SNPs/genome/year [53,290,313]. Therefore, MGIs, and in particular MSIs, could have implications during infections caused by slow-growing pathogens by exposing the host to genetically and phenotypically distinct strains at the same time [156,172,304]. Blood antigens or other markers that could be used to accurately predict (or diagnose) JD progression are still lacking [156,294], which could also be a consequence of MGIs. In the case of

other diseases, MGIs are known to interfere with the host immune response, possibly due to antigenic differences between the infecting strains [294,304,305].

5.3 Objectives

While MAP MGIs have been reported in dairy animals using multi-locus SSR (ML-SSR) [315] and SNP [116,289] based analysis, a focused study on the topic is lacking. To conduct an in-depth investigation into the presence of MAP MGIs in dairy cattle, WGS was used to examine genetic variations in multiple MAP isolates derived from animals shedding high levels of the bacterium. This study identified unique ML-SSR and SNP patterns from isolates both within the same animal and between different animals, allowing for the identification of MGIs, and in some cases, providing evidence for either microevolution or MSI events.

5.4 Results and Discussion

Information on strain dynamics in animals infected with different MAP variants, whether it be due to MSIs or microevolution (collectively referred to as MGIs), is scarce in the current literature. Therefore, multiple MAP isolates from single infected animals were obtained for in-depth genomics analysis. As part of a larger ongoing project in 2013-2017, 2-10.5-year-old JD-positive dairy animals from herds located in the provinces of QC and ON in Canada were examined. These animals (n=555) served as the source of the blood and fecal samples that were collected on a semi-annual basis and were used in the described analysis.

5.4.1 Host JD status and MAP strain isolation

Serum ELISA and fecal qPCR-based analysis of collected samples were performed as previously described in Chapter IV of this thesis. Serum ELISA measures the presence of specific anti-MAP antibodies which indicates whether the host has recognized the pathogen and mounted a humoral response against MAP. The fecal qPCR assay quantifies the amount of MAP

DNA present in fecal samples, while the fecal decontamination and culture protocol confirmed the presence of live MAP. To examine the presence of MGIs, 14 high-shedding animals from four farms (three from QC and one from ON), were selected for further analysis **(Appendix Figure A3)**. The last fecal sample collected from each cow was cultured on solid media, confirming the high level of MAP shedding status as described previously [178]. Isolation of 10 MAP colonies was performed from each animal, except for animal A20, where only nine colonies could be obtained **(Table 2.4)**. Each colony was individually grown in liquid media (n=139 axenic cultures) and used for isolating bacterial DNA for whole genome sequencing using Illumina NovaSeq 6000 technology, just as performed in Chapter IV.

5.4.2 SNP-based Phylogenetic Analysis

After assembling the sequence data, the 139 genome sequences were qualified as being high-quality assemblies, with an average coverage of 14X, and approached near completeness (99.28%) **(Appendix Tables A10** and **A11)** [361]. Based on a selection of genomic markers [272,273,313,614], all 139 isolates from the current study were classified as type II strains. To identify core genome SNPs for constructing phylogenies for the 139 isolates, MAP K-10, a prototypical type II strain associated with dairy cattle [112,638], was used as a reference sequence **(Figure 5.1)**. Detailed analysis showed that the number of SNP variants detected across all isolates was low (67-101 SNPs), but the diversity increased to 125-158 variants when insertions and deletions were taken into consideration **(Table 5.1)**.

Based on the constructed core SNP phylogeny, MAP isolates from the current study broadly clustered into four clades, where isolates derived from the same animal often clustered closely together **(Figure 5.1)**. However, this was not the case for four animals (A25, A34, A36, and AM3), which had MAP isolates belonging to different clades **(Figure 5.1)**, suggesting that

these animals were infected with phylogenetically distinguishable strains, and could be classified as having MSIs. Core genomes of MAP isolates derived from the same animal were directly compared to examine this further **(Figure 5.2** and **Appendix Figure A4**). While pairwise comparisons of isolates from some animals showed that they were closely related **(Figure 5.2 A)**, MAP isolates from animals A25, A34, A36, and AM3 showed large SNP differences, which confirms the presence of MSIs caused by substantially diverse isolates that could not have arisen by microevolution **(Figure 5.1)**. Examination of MAP isolates from animal A25 showed that all isolates cluster together, except for isolate 26 **(Figure 5.1)**, which shows a SNP difference of 37 compared to the next closest related strain **(Figure 5.2 B)**. Similarly, isolate 128 from animal A34 does not cluster with other isolates from this animal **(Figure 5.1)** and differed by 66 SNPs **(Figure 5.2 C)**. MAP isolates from animal A36 are divided into two clades, with eight clustering together, and the remaining two isolates (251 and 253) clustering separately **(Figure 5.1)**. A pairwise comparison of the closest related isolates between these two clades (251 and 254) showed a difference of 62 SNPs **(Figure 5.2 D)**. The 10 isolates from animal AM3 clustered evenly into two different clades **(Figure 5.1)**, where isolates 241 and 242 from the two clades differ by 69 SNPs **(Figure 5.2 E).** The differences in clustering patterns **(Figure 5.1)** and unique SNP numbers **(Figure 5.2)** of MAP isolates from these four animals strongly suggest instances of MSIs, indicating independent co-infection with different isolates.

5.4.3 SNP evolution rates and MSIs

The distribution of MAP isolates from the same animal into different clades of the phylogenetic tree provides clear evidence in support of MSI events **(Figure 5.1)**, but based on the data, it is possible to examine strain evolution trajectories at a finer level. The genome evolution rate of MAP is very slow, with studies suggesting that changes occur in the range of

0.1-0.5 SNPs/genome/year [53,290,313]. The animals included in this study were between 4-10.5 years of age at the time of sample collection. Working under the assumption that animals were infected shortly after birth and using a conservative genome evolution rate of 0.5 SNPs/genome/year, the theoretical number of SNPs that would be observed in the sequenced isolates if they evolved from a strain that caused the initial infection can be calculated **(Table 5.2)**. By performing SNP analysis on isolates that grouped closely together in phylogenetic trees constructed on a per-animal basis **(Figure 5.2** and **Appendix Figure A4**), it could be determined if the number of observed SNPs in each pairwise comparison matched theoretical calculations required to provide support for intra-host microevolution or potential MSI events.

Based on expected differences in MAP SNP frequencies, within-animal comparisons between isolates provided evidence for both microevolution and MSI events in all but two of the 14 animals examined **(Table 5.3)**. MAP isolates from animals A32 and AM2 only showed evidence for potential microevolution, but not MSI events, as SNP thresholds for all pairwise comparisons were within the expected range for microevolution-based changes, a threshold of three SNPs in both cases **(Table 5.2)**. As previously mentioned, four animals showed clear indications of MSI events due to the presence of isolates from distinct clades **(Figures 5.1** and **5.2)**. The evidence provided by examining SNP patterns in pairwise comparisons supports these results and identifies potential MSIs at a higher resolution compared to other genotyping methods [53,290,313]. Although these theoretical SNP calculations suggest clear differentiation between microevolution and MSIs, it cannot be completely ruled out if closely related MAP isolates with SNP differences below the assigned thresholds can infect animals through an MSI event instead of developing through microevolution. In some instances, MAP isolates with

identical SNPs had different indels; due to which clonal isolates were not identified in the current study.

5.4.4 *In-silico* analysis of repetitive DNA elements

Previous studies have indicated that MAP has a closed pangenome [57,288,616], implying that horizontal gene transfer does not significantly contribute to genome evolution in the bacterium [639,640]. The low SNP evolution rate and apparent lack of horizontal gene transfer in MAP make unstable repetitive elements both a driving force for evolution and attractive targets for strain differentiation using molecular methods [641,642]. Both MIRU-VNTR and ML-SSR are significant sources of genetic polymorphisms in mycobacteria and are often used in strain typing [642–644]. *In-silico* analysis of eight MIRU-VNTR loci, specifically MIRUs 292 and X3 along with VNTRs 25, 47, 3, 7, 10, and 32 [287], and 11 SSR loci [286] was performed on the 139 isolates from the current study **(Table 2.4)** to determine how they compare with the SNP based phylogenies of the respective isolates.

Of the eight MIRU-VNTR loci analyzed, none showed variability in the 139 strains examined (DI=0), sharing only a single INMV-MIRU pattern [373], INMV 2 **(Table 2.4)**. These results agree with prior observations described in Chapter IV, where INMV 2 was also the only INMV-MIRU pattern observed in MAP isolates from the same four herds. Complete ML-SSR profiles were also obtained for 127 isolates and were more informative than MIRU-VNTR markers **(Figure 5.3** and **Table 2.4)**, an observation that has been noted previously [627,628]. The ML-SSR profiles of 12 of the 139 isolates could not be validated and were therefore not used in the analysis. Of the 11 SSR loci examined, SSR1 (n=128), SSR2 (n=138), and SSR7 (n=139) showed variations, whereas the remaining eight loci had identical repeat patterns across all isolates **(Figure 5.3** and **Table 2.4**). SSR1 has the highest DI value (0.9081), followed by SSR2

(0.8051), then SSR7 (0.3241), which agrees with the previous observations described in Chapter IV. For SSR1 and SSR2, fifteen and nine distinct repeat pattern numbers were identified, respectively, whereas SSR7 only showed two different patterns **(Table 2.4)**.

For the 127 isolates for which unambiguous SSR sizes could be assigned, a comparison of the results to the INMV-SSR database revealed seven matches along with 19 previously unreported ML-SSR patterns **(Table 5.4)** [373]. Two of these novel patterns (named N/P 5 and N/P 16) were previously described in Chapter IV (as N/P 12a and N/P 15, respectively), whereas the remaining 17 have not been previously reported within either the literature or available collections. As previously mentioned in Chapter IV, an important point of consideration is that the MAC-INMV database does not individually classify mononucleotide SSRs greater than 11 nucleotides, but instead groups them (as "+"). This is normally due to the limitations of conventional DNA sequencing technologies to provide accurate sequences for such repeats [645]. In both this chapter and Chapter IV, NovaSeq technology was used [646], which can provide sequence data for longer repeats leading to new ML-SSR types that would otherwise be categorized as "+". Four of the INMV patterns identified in this study (ML-SSR 10, ML-SSR 7, ML-SSR 8, and ML-SSR 93) could be classified further using this increased resolution **(Table 5.4)**, indicating greater MAP diversity than otherwise suggested by the MAC-INMV database. When accounting for the increased resolution, the total number of unique patterns and DI value increased from 9 to 71 and 0.7942 to 0.9839, respectively. All 14 animals examined in this study contained distinct MAP isolates with different ML-SSR patterns, with a minimum of five and a maximum of ten patterns found within MAP from each animal **(Figures 5.1** and **5.3)**. Of the 71 ML-SSR patterns found in this study, many were shared by MAP isolates distributed between he four major clades identified based on SNP profiling **(Figure 5.1).**

Examination of certain MAP isolates from the same animal with identical SNPs, indels, and MIRU-VNTR loci, could still be easily differentiated due to variations in SSR1 and SSR2 **(Table 5.5)**. These findings agree with recent observations made while analyzing single MAP isolates from infected animals, suggesting that ML-SSR typing, like MIRU-VNTR [630], is subject to homoplasy, and does not directly correlate with SNP-based phylogenies, Previous studies have shown that microsatellite loci within bacterial genomes, such as the 11 SSR repeats examined here, have a faster rate of evolution compared to the rest of the genome [641,647]. The presence of multiple isolates that are only different at these SSR loci, but are otherwise identical, suggests that the homopolymeric repeats found at SSR1 and SSR2 evolve at a faster rate than the rest of the MAP genome. Therefore, while SSRs might not be good targets for studying phylogenetic lineages due to homoplasy, their differential evolutionary rates as compared to the rest of the genome could have significant functional implications.

5.4.5 SSR1 and SSR2 variation and distribution in diverse MAP isolates

Following the observation that MAP homopolymeric repeats associated with SSR1 and SSR2 are highly variable and homoplastic, it became necessary to examine the diversity and distribution of these repeat sequences within the genome sequences of all available MAP isolates, including those acquired from previous sequencing work (n=192) and those of comparable quality and coverage present in the public SRA and NCBI Nucleotide databases (n=1429) **(Appendix Table A1)**. Of the 1621 MAP genomes examined, 182 and 53 sequences for SSR1 and SSR2, respectively, had to be discarded due to ambiguous coverage of the specific genomic loci **(Appendix Table A1)**. For the remaining records, the repeat sizes (as nucleotides or nt) at both loci were recorded.

SSR1 and SSR2 are located in open reading frames that are designated within this study as *ORF1* (K-10 locus tag MAP_RS08325) and *ORF2* (K-10 locus tag MAP_RS23340), respectively, in the annotated MAP K-10 genome [112,638]. The DNA sequences of *ORF1* and *ORF2* in the reported MAP K-10 genome sequence contain 19 and 10 repeats for SSR1 and SSR2, respectively, and were assigned as Reading Frame-1 (RF-1) in this study **(Appendix Table A1)**. Based on this assignment, different SSR1 and SSR2 repeat sizes were classified into three different reading frames (RF-1, RF-2, or RF-3) depending on how they affected the predicted amino acid sequence of the cognate gene product **(Figures 5.4** and **5.5**, and **Table 5.6)**. SSR1 showed a total of 1107, 156, and 182 records which were categorized as RF-1, RF-2 and RF-3, respectively **(Figure 5.4 A)**. In comparison, SSR2 had a total of 706 records classified as RF-1, 369 as RF-2, and 493 as RF-3 **(Figure 5.4 B)**. Therefore, it appears that there is some bias in SSR length selection at both loci to maintain the respective ORFs in RF-1, which seems to be prominent for ORF1 **(Figure 5.4** and **Appendix Table A1)**. The prevalence of the most common mononucleotide repeat lengths for SSR1 and SSR2 were examined for each reading frame to determine if there were any preferences **(Figures 5.4** and **5.5 A)**. The most common repeat sizes for the reading frames in SSR1 were 7-nt (RF-1, 88.71% of repeats), 8-nt (RF-2, 32.69% of repeats), and 9-nt (RF-3, 39.77% of repeats) **(Figure 5.4 A)**, while the most common repeat sizes for the reading frames in SSR2 were 10-nt (RF-1, 86.12% of repeats), 11-nt (RF-2, 87.53% of repeats) and 9-nt (RF-3, 61.46%) **(Figure 5.4 B)**.

The observed distribution of *ORF1* and *ORF2* reading frames and repeat numbers for SSR1 and SSR2 were found to be statistically significant according to a χ^2 goodness of fit test, suggesting that they were not distributed equally among the population **(Figures 5.4** and **5.5)**. Although there seems to be some preference for SSR1 and SSR2 lengths, variations result in the

shifting of the entire reading frame of the respective genes containing them. Several bacterial and fungal species have mechanisms that allow for phenotypic switching [648,649], enabling them to reversibly alter select phenotypes. One such mechanism used for phenotypic switching is slipped strand mispairing (SSM) during DNA replication, which is especially prevalent in SSR regions, making them prone to insertions and deletions [647,650,651]. Therefore, the presence of hypervariable homoplastic repeats in *ORF1* and *ORF2* in the otherwise slowly evolving MAP genome, suggests that SSR1 and SSR2 provide a mechanism to promote genetic changes to alter the functions of the respective genes.

5.4.6 Predicted influence of SSR1 and SSR2 on genes containing them in MAP

The predicted ORF1 product (WP_010949291.1) from MAP K-10 (RF-1) is comprised of 207 amino acids, where the contained SSR1 encodes for a series of glycines beginning at residue 107 **(Figure 5.5 A)**. In addition, all three versions of ORF1 (RF-1, RF-2, and RF-3) contain Sec-dependent secretion signals at their N-termini [379,652], with signatures that suggest signal peptidase II (SPII) dependent processing and membrane attachment [653–655] **(Figure 5.6)**. Therefore, all ORF1 variants are predicted to be surface-exposed proteins in MAP and are homologous (80-100% identity and coverage) to phosphatidylethanolamine-binding (PEB) domain (COG1881) containing mycobacterial hypothetical proteins listed as YbhB/YbcL family Raf-kinase inhibitor-like proteins. Further examination of *ORF1* confirmed that the gene is only found in mycobacteria, specifically members of the MAC, including *M. avium* subsp. *lepraemurium*, which has also been classified as a member of the complex recently **(Appendix Table A12)** [52,77]. In addition, *ORF1* orthologues from all MAC members match the RF-1 homologue, with no evidence of the variable region observed in MAP **(Figure 5.5 A)**. Members from the PEB protein (PEBP) superfamily (also referred to as Raf-kinase inhibitor proteins, RKIP) are found in diverse

organisms where they serve a variety of roles, including regulating Raf/MAP kinase and NF-κB signalling pathways in mammals [656–658], which in turn control immunity, stress responses, apoptosis, and differentiation. In bacteria, PEPB/RKIP proteins from the YhbB/YbcL family have proposed roles in the phosphorylation-based regulation of signalling pathways [659], and YbcL has been noted to contribute to immune modulation in uropathogenic *Escherichia coli* [660]. While many PEBP/RKIP proteins only have intracellular functions, some PEBP proteins, such as mammal protein PEBP4 (and its various homologues), can also be secreted [661,662].

In-silico examination of the three *ORF1* variants expected to be produced due to reading frame shifts based on different SSR1 lengths indicates that the first 106 amino acid residues, including the Sec-signal sequence, are conserved in all of them **(Figure 5.5A)**. The SSR1 sequence encodes a series of glycine residues (starting at residue number 107), the number of which changes depending on the length of the repeat present. Different SSR1 lengths alter the reading frame, leading to the production of proteins with different sizes and amino acid sequences in the region following the glycine repeat **(Figure 5.5A)**. In addition, the predicted PEBP domain of all three ORF1 variants begins at residue 39 and varies in size depending on the protein product. Differences were also observed in the predicted structures of the three ORF1 variants obtained using Alphafold2 **(Figure 5.5C)**, which was implemented using Colabfold [380]. A high-confidence predicted structure for the C-terminal region of ORF1 could be obtained for RF-1, which was not the case for RF-2 and RF-3. This drop in proteins structure confidence suggests that the latter two variants adopt structures that are significantly different and cannot be confidently predicted based on currently available templates and algorithms [663–665]. Due to evidence suggesting the export of ORF1 through the Sec/SPII pathway, this would allow for the expression of the predicted PEBP domain and subsequent region on the extracellular surface,

where the observed variability between reading frames may have an impact on antigenic variation or other host-pathogen interactions [666,667].

In comparison, the predicted ORF2 product from MAP K-10 (WP_134797017.1) comprises 91 amino acids, where SSR2 encodes a series of glycines starting at residue 56. All ORF2 variants arising from the three different reading frames are predicted to be cytoplasmic, and hypothetical proteins similar to ORF2 (with 80-100% identity and coverage) are only present in MAC members. In addition, the ORF2 orthologue from *M. avium* subsp. *avium*, *M. avium* subsp. *hominissuis* and *M. avium* subsp. *lepraemurium* sequences also contain the SSR2 homopolymer sequence. Only a single *M. avium* subsp. *lepraemurium ORF2* sequence (CP021238.1) where SSR2 comprised of 9-nt, was identified during homology searches **(Appendix Table A12)**. In the reported *M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis* sequences, SSR2 ranges from 9-11 nt, suggesting the production of protein variants in MAP, *M avium* subsp. *avium* and *M. avium* subsp. *hominissuis* due to frame shifting **(Appendix Table A12)**. As with ORF1, the impact of reading frame shifts caused by SSR variation in MAP ORF2 products was also examined. The first 55 residues preceding the glycine repeat are identical in all three ORF2 variants, which otherwise differ in their C-terminal sequences and overall sizes **(Figure 5.5 B)**. The impact of these variations is highlighted in the predicted AlphaFold2 generated high-confidence structures, suggesting that ORF2 RF-1 and RF-2 comprise two alpha helices, while RF-3, only contains one **(Figure 5.5 D)**, which in theory should lead to major alterations in cognate protein activity.

5.5 Conclusion

By analyzing the genome sequences of MAP isolates from the same animal, the identification of MGIs and the ability to distinguish between MSIs and microevolution-based

events was possible. MGIs could also be classified into different categories based on the genetic relatedness of MAP isolates, ranging from those caused by MAP with diverse SNP-based phylogenies to one where the isolates were more closely related to each other. It is important to note that while distinction between microevolution and MSI events can likely be determined using SNP-based calculations, the proposed stable nature of the MAP genome warrants further investigation on the *in vivo* contributions of the two processes. From the results in this Chapter, it was noted that SSRs allowed for enhanced discrimination between MAP isolates when complemented with SNP analysis, despite the homoplastic nature of the former, and both methods were more discriminatory than MIRU-VNTR. Maximum diversity was observed in the case of SSR1 and SSR2, and in some instances, they were the only sequences that differed between otherwise isogenic MAP isolates **(Figure 5.2** and **Table 5.5)**. It was also noted that all *ORF1* sequences from other MAC subspecies corresponded to RF-1, almost all of which contained the seven-nucleotide "GGCGGGG" sequence instead of the "G" homopolymer present in MAP. *M. avium* subsp. *lepraemurium* had the most dissimilar sequence, which corresponded to "AGCGGGG". This analysis showed that the most common SSR1 repeat size in MAP was seven **(Figures 5.3**, **5.4** and **Appendix Table A1)**, resulting in a protein sequence that matches the predicted *ORF1* and *ORF2* sequences [67,633,634]. As mentioned previously, the *M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis* orthologous *ORF1* contains a "GGCGGGG" sequence, raising the possibility that a C→G transversion in the sequence of an ancestral MAC member led to the establishment of the unstable SSR1 mononucleotide homopolymer found in MAP [668,669]. This, in turn, resulted in the variety of repeat sizes reported for SSR1 in MAP, while also affecting the reading frame of *ORF1* **(Appendix Table A12)**. In the case of *ORF2*, the *M. avium* subsp. *avium*, *M. avium* subsp. *hominissuis* and *M. avium* subsp. *lepraemurium* orthologs also

contain the SSR2 homopolymer, where the *M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis* sequences exhibit variations similar to those observed in MAP. As with MAP, this suggests that *M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis* may have a similar capacity to produce different gene products using alternate reading frames.

In general, variations caused in the amino acid sequences and corresponding structures of MAP proteins due to the presence of variable SSRs in cognate genes have not received much attention. The ORF1 and ORF2 variants predicted in the current study reinforce the importance of variable genetic elements in affecting protein products, beyond just being used as molecular targets in strain typing studies. While the predictions described using *in-silico* protein analysis reveals novel information, not much is known about the functions of ORF1 and ORF2, beyond that ORF1 is a predicted surface-exposed protein possibly involved in modulating host responses based on its homology to other known proteins [659,670]. It can be hypothesized that changes in the cognate *ORF1* product caused by SSR-associated frame shifting could potentially affect the interaction of MAP with the host, a point that may be tested in future work. It is currently not known how these SSRs change or evolve in infecting MAP isolates, but regardless of their origin, MAP isolates with different SSRs from the same animal can be categorized as MGIs for now, as they are technically not isogenic.

There have been some studies on co-infections involving unrelated species of pathogens in humans and animals [671–674], and viral MGIs have also been documented [675–678]. In comparison, not much is known about the implications of bacterial MGIs, but such phenomena have been modelled or studied using other unicellular eukaryotic pathogens [679]. For example, MSIs in mice caused by either *Plasmodium chabaudi* [680] or *Trypanosoma brucei* [681], the causative agents of rodent malaria and human African sleeping sickness, respectively, can lead to mutual suppression

or decreased disease severity during the acute phase of infection. In comparison, during the chronic phase of infection, competition between different pathogen strains is low and MSIs are harder to control by the host immune system as compared to clonal infections caused by a single strain [676]. Since JD is characterized by having a long subclinical phase of infection with unpredictable disease progress, it remains to be determined if an interplay by different MAP genotypes could have similar effects on either pathogen clearance or disease progression to full-blown clinical JD. The fact that MAP can survive in the environment, multiple strains can infect animals, and the possibility that the pathogen can further evolve within the host due to unstable genetic elements, could have major implications on the spectrum of clinical outcomes and the ability to control JD.

5.6 Figures and Tables

5.6.1 Figures

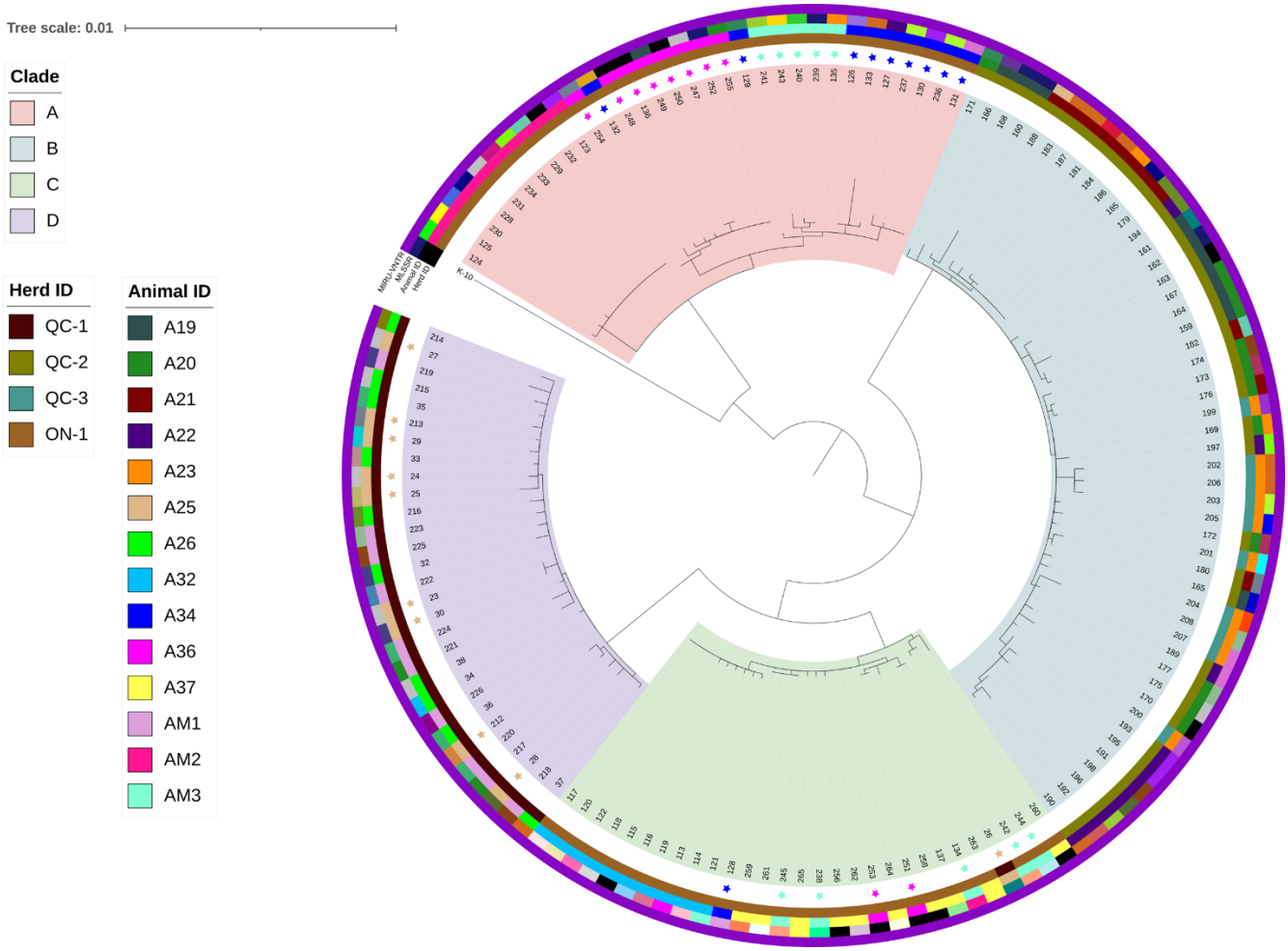


Figure 5.1: Core SNP-based phylogeny of 139 MAP isolates obtained from the feces of 14 animals examined. Isolates are distributed among four major clades based on SNP profiles, where each clade has been coloured differently as indicated. For each isolate the concentric circles indicate (from inside to outside): the herd where the animal came from (ID), the animal from which the isolate originated (ID), the ML-SSR pattern and the INMV MIRU-VNTR type. Colours used for depicting herd and animal IDs are included. ML-SSR patterns labelled in black are unknown, while all other colours represent individual strain types (details are provided in **Table 2.4**). All isolates in this study showed the same MIRU-VNTR pattern (INMV 2) as indicated by a single colour. The differently coloured stars indicate isolates from herds that are distributed among more than one of these four major clades, with the colour of the star reflecting the animal the isolate was derived from.

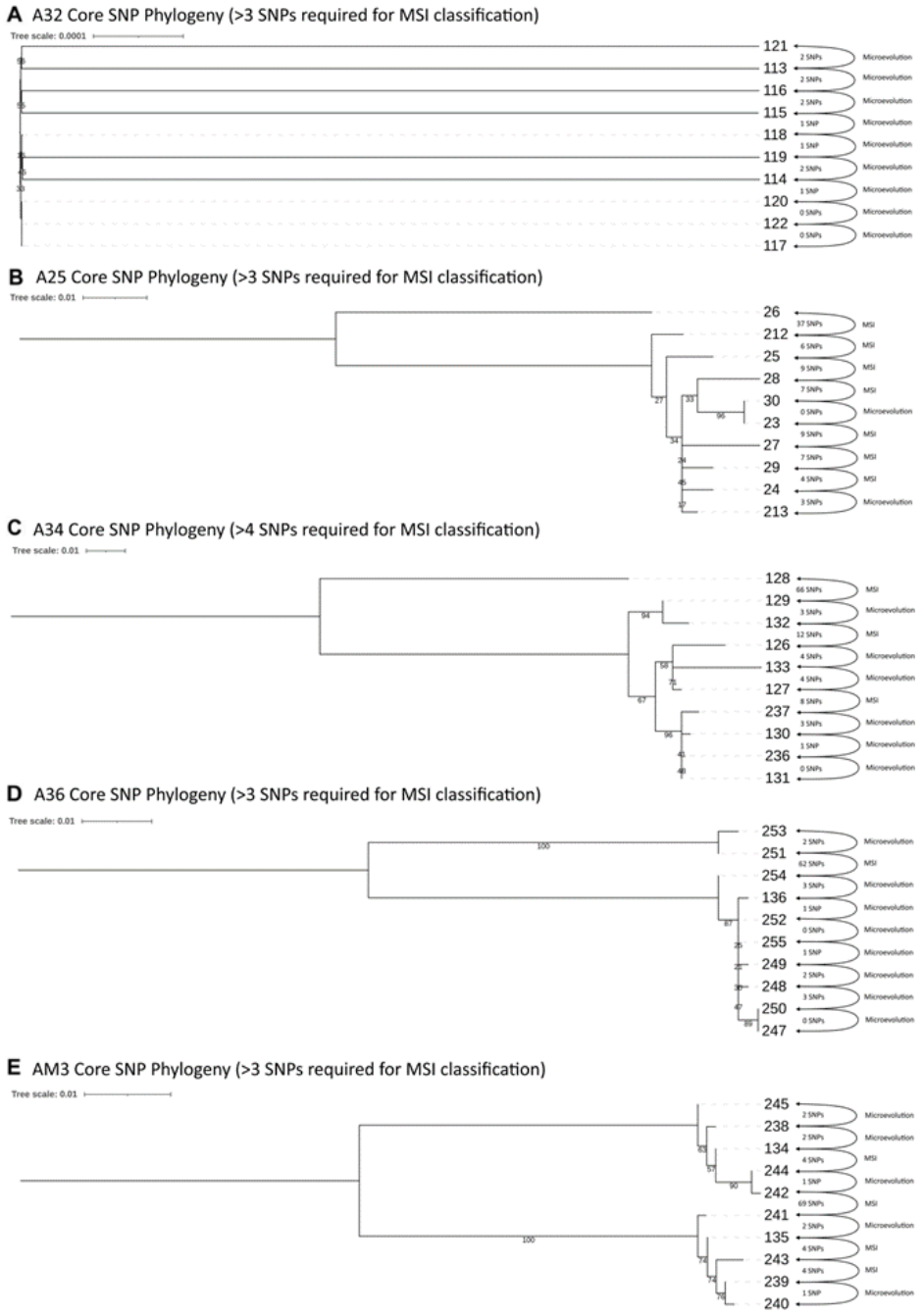


Figure 5.2: Core SNP phylogenetic trees of select MAP isolates constructed on a per-animal basis. SNPs in all isolates from each animal (n=10, except A20, where n=9) were used for tree construction. The arrows between related MAP pairs derived from the same animal indicate the number of core SNPs that are different between them, and whether this correlates with microevolution or MSI. **(A)** Core SNP phylogeny of MAP isolates from animal A32 as an example where all isolates are closely related, and any SNP differences can be explained on microevolution. **(B-E)** Individual phylogenies of MAP isolates from animals, A25, A34, A36 and AM3, respectively, were distributed across more than one clade in **Figure 5.1**. Isolates from these four animals had much higher SNP differences, alluding to more instances of MSIs.

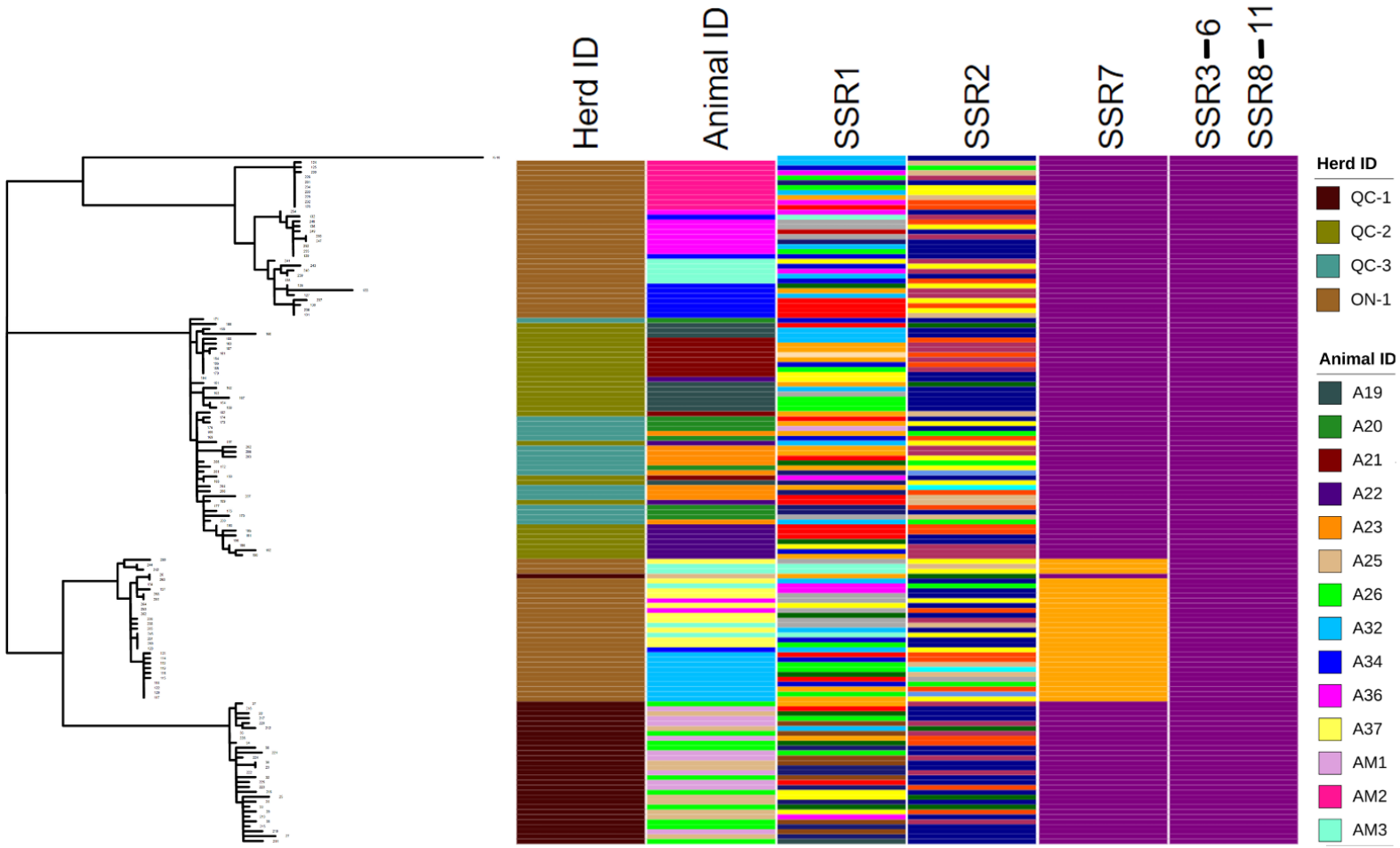


Figure 5.3: Visual depiction of the SSR diversity observed in MAP isolates. The phylogenetic tree shown is identical to the one in **Figure 5.1**. Each column represents SSR profiles at a specific locus (1-11), where each line indicates the SSR in a MAP isolate, which is colour coded according to repeat length. The SSR loci that did not show any diversity (were the same in all isolates: SSR3, 4, 5, 6, 8, 9, 10 and 11) are indicated by the last column on the right for the sake of simplicity.

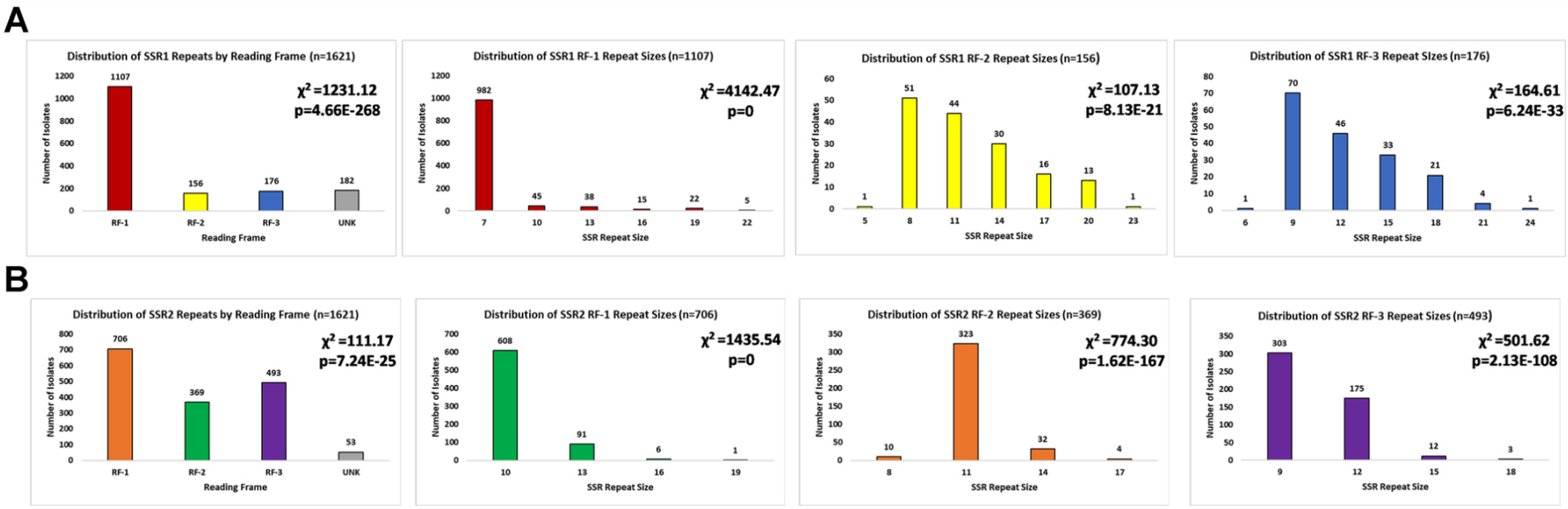


Figure 5.4: Distribution of (A) SSR1 and (B) SSR2 repeat sizes in genome sequences of MAP isolates from the current study and those available in public databases. In MAP, SSR1 and SSR2 are situated in ORF1 and ORF2, respectively, and the sequences of the genes in the published MAP K-10 genome sequence were assigned reading frame one (RF-1). Based on this, any alteration in the copy number of SSR1 or SSR2 leads to reading frameshifting in the respective ORFs (designated as RF-2 and RF-3). (A and B) The first panel (from the left) shows the distribution of repeat sizes among the three possible reading frames for each ORF, whereas the following panels show the distribution of repeat sizes within RF-1, RF-2 and RF-3 for the respective ORFs. The number of genome sequences used in each analysis is included (n) along with the $\chi 2$ test of independence and p values obtained.

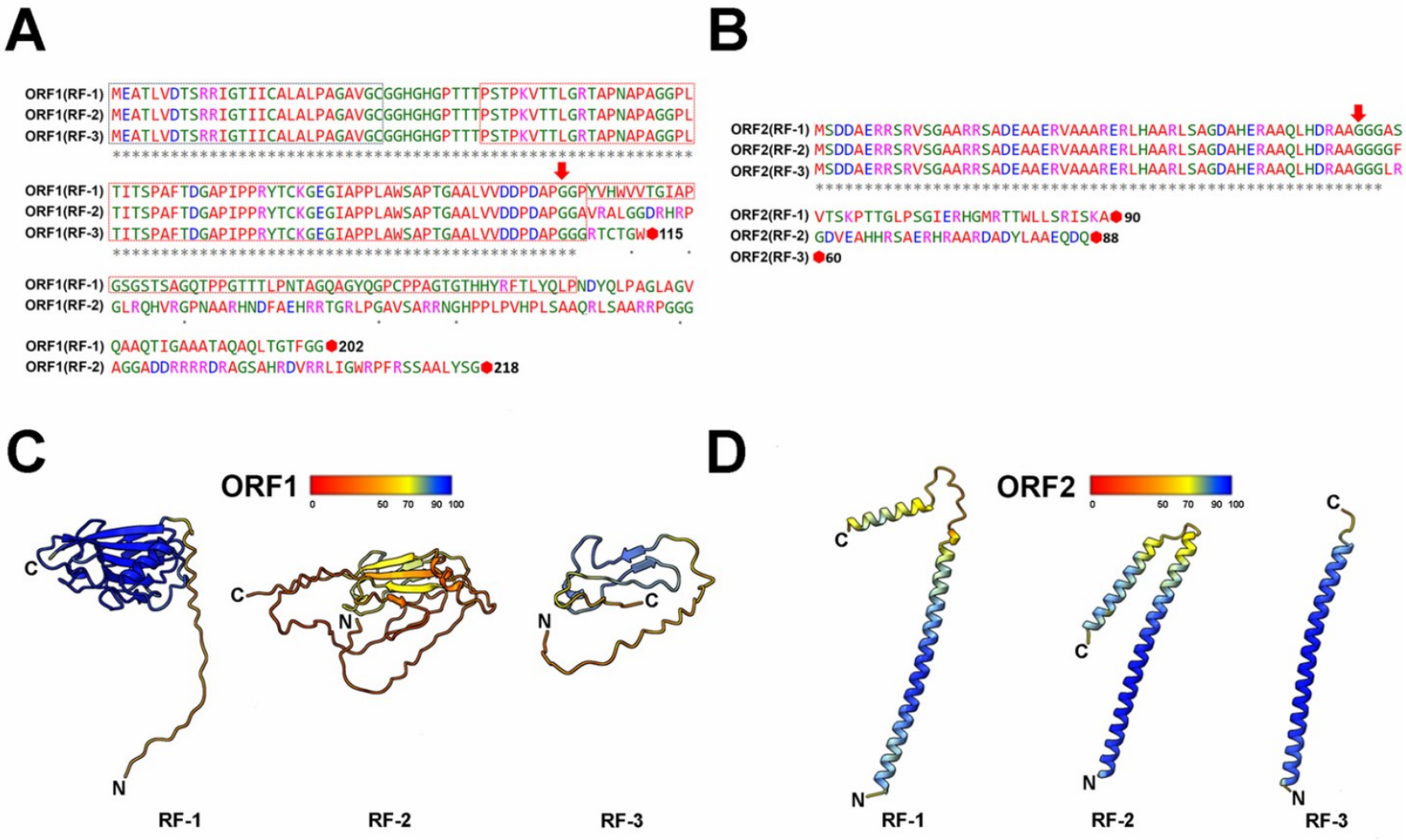


Figure 5.5: Predicted impacts of reading frame shifts caused by SSR1 and SSR2 variations on the putative ORF1 and ORF2 proteins, respectively. Protein alignments of (A) ORF1 and (B) ORF2 show products from each reading frame (RF-1 RF-2 and RF-3), respectively. The most common SSR repeat sizes for each reading frame (as indicated in **Figure 5.4**) were used to obtain the amino acid sequences of the proteins shown. (A and B) The red arrows indicate the glycine repeat encoded by the

variable SSRs. For each ORF, amino acids that are identical across all three sequences are indicated by (*) while those conserved in two of the three variants in each alignment are indicated by (●). Stop codons (red hexagons) and the number of residues comprising each protein are also indicated. (A) The gray box indicates the Sec II signal peptide sequence in the three ORF1 variants, while the red box represents the phosphatidylethanolamine-binding (PEBP) domain identified in each sequence. Predicted protein structures of (C) ORF1 and (D) ORF2 were obtained using Alphafold2. (C and D) The computed structures of protein variants corresponding to the three reading frames (RF-1, RF-2 and RF-3) are shown, and their N and C termini are indicated. Colouration is based on the predicted local distance difference test (pLDDT) score for confidence in the structures, where blue indicates high confidence and red regions low, as indicated by the scale.

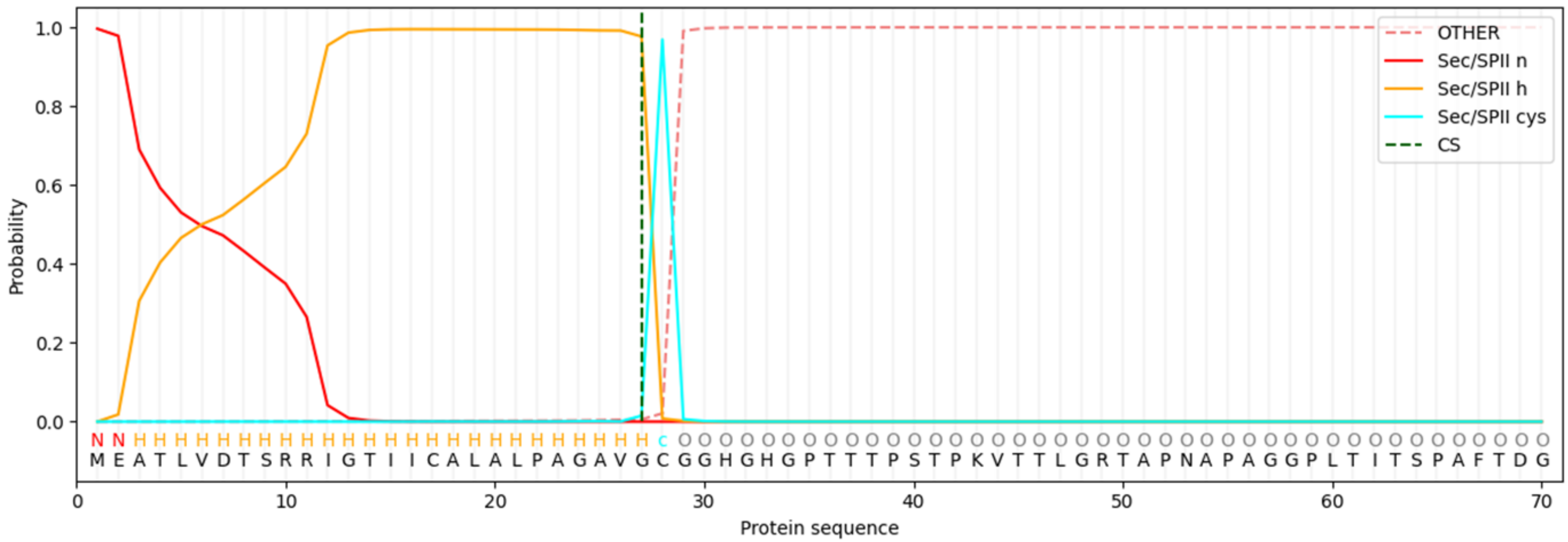


Figure 5.6: SignalP 6.0 analysis result for the ORF-1 RF-1 amino acid sequence. Results indicate that the signal peptide is transported by the Sec translocon and is cleaved by Signal Peptidase II. The letters above the protein sequence and in the caption are indicative of the N-terminal region (N), the center hydrophobic region (H), and the conserved cysteine in the +1 cleavage site (CS) of lipoproteins (c) found within the signal peptide. Regions labelled “O” are not part of the signal peptide sequence.

5.6.2 Tables

Table 5.1: Snippy Results of MAP strains (n=139) using type II strain K-10 as a reference sequence

Isolate Number	Farm ID	Animal Number	SNPs	MNPs	INS	DEL	COMPLEX	TOTAL
#159	QC-2	A19	91	2	9	42	5	149
#160	QC-2	A19	94	1	9	42	7	153
#161	QC-2	A19	93	0	9	43	7	152
#162	QC-2	A19	93	2	9	44	5	153
#163	QC-2	A19	93	2	9	42	5	151
#164	QC-2	A19	91	1	9	42	6	149
#165	QC-2	A19	92	2	13	42	5	154
#166	QC-2	A19	95	1	9	43	6	154
#167	QC-2	A19	93	1	9	45	5	153
#168	QC-2	A19	94	1	9	44	6	154
#169	QC-3	A20	92	1	10	42	6	151
#170	QC-3	A20	95	2	9	42	5	153
#171	QC-3	A20	91	1	9	44	6	151
#172	QC-3	A20	94	1	11	43	6	155
#173	QC-3	A20	94	1	10	42	6	153
#174	QC-3	A20	94	2	10	43	5	154
#175	QC-3	A20	93	1	10	43	6	153
#176	QC-3	A20	92	0	10	42	7	151
#177	QC-3	A20	91	2	11	44	5	153
#179	QC-2	A21	92	1	9	43	6	151
#180	QC-2	A21	93	2	12	42	5	154
#181	QC-2	A21	91	0	10	43	7	151
#182	QC-2	A21	94	1	11	45	6	157

#183	QC-2	A21	94	2	9	45	5	155
#184	QC-2	A21	93	1	9	42	6	151
#185	QC-2	A21	93	2	10	43	5	153
#186	QC-2	A21	93	0	10	43	7	153
#187	QC-2	A21	93	1	9	43	6	152
#188	QC-2	A21	95	2	10	44	5	156
#189	QC-2	A22	91	1	10	44	6	152
#190	QC-2	A22	95	2	10	42	5	154
#191	QC-2	A22	96	1	10	43	6	156
#192	QC-2	A22	92	1	9	43	7	152
#193	QC-2	A22	93	2	10	42	5	152
#194	QC-2	A22	91	0	9	43	7	150
#195	QC-2	A22	95	2	10	42	5	154
#196	QC-2	A22	95	2	10	43	5	155
#197	QC-2	A22	95	2	11	43	5	156
#198	QC-2	A22	93	2	0	45	6	155
#199	QC-3	A23	92	2	9	45	5	153
#200	QC-3	A23	92	2	10	46	5	155
#201	QC-3	A23	92	1	10	45	6	154
#202	QC-3	A23	96	1	10	44	6	157
#203	QC-3	A23	98	2	10	43	5	158
#204	QC-3	A23	93	2	10	45	6	156
#205	QC-3	A23	93	2	10	42	5	152
#206	QC-3	A23	98	1	10	42	6	157
#207	QC-3	A23	95	2	9	45	6	157
#208	QC-3	A23	94	0	10	45	7	156
#212	QC-1	A25	96	3	7	40	4	150
#213	QC-1	A25	97	2	7	40	5	151

#23	QC-1	A25	100	3	7	43	4	157
#24	QC-1	A25	98	3	8	41	4	154
#25	QC-1	A25	98	3	7	42	4	154
#26	QC-1	A25	82	1	10	40	6	139
#27	QC-1	A25	101	2	8	39	5	155
#28	QC-1	A25	99	1	7	41	6	154
#29	QC-1	A25	97	1	7	39	6	150
#30	QC-1	A25	100	1	7	40	6	154
#214	QC-1	A26	97	3	7	44	4	155
#215	QC-1	A26	97	0	8	43	7	155
#216	QC-1	A26	98	2	8	41	5	154
#32	QC-1	A26	100	2	8	43	5	158
#33	QC-1	A26	97	2	8	43	5	155
#34	QC-1	A26	97	2	7	40	5	151
#35	QC-1	A26	98	3	9	40	4	154
#36	QC-1	A26	96	1	7	40	6	150
#37	QC-1	A26	97	2	6	40	6	151
#38	QC-1	A26	99	1	6	39	7	152
#113	ON-1	A32	86	1	9	40	5	141
#114	ON-1	A32	84	2	9	40	4	139
#115	ON-1	A32	87	1	9	40	5	142
#116	ON-1	A32	87	2	9	40	4	142
#117	ON-1	A32	86	1	9	41	5	142
#118	ON-1	A32	86	1	9	40	5	141
#119	ON-1	A32	87	1	9	43	5	145
#120	ON-1	A32	86	2	9	40	4	141
#121	ON-1	A32	84	2	8	40	5	139
#122	ON-1	A32	86	0	9	41	6	142

#126	ON-1	A34	70	1	8	43	7	129
#127	ON-1	A34	69	2	9	43	5	128
#128	ON-1	A34	83	1	9	40	5	138
#129	ON-1	A34	69	1	9	42	6	127
#130	ON-1	A34	73	0	10	43	7	133
#131	ON-1	A34	72	1	9	44	6	132
#132	ON-1	A34	71	1	10	42	6	130
#133	ON-1	A34	75	0	9	42	8	134
#236	ON-1	A34	72	1	10	43	6	132
#237	ON-1	A34	72	1	9	43	6	131
#136	ON-1	A36	71	3	9	42	4	129
#247	ON-1	A36	72	0	9	42	7	130
#248	ON-1	A36	71	3	9	42	4	129
#249	ON-1	A36	71	1	9	43	6	130
#250	ON-1	A36	72	3	9	44	4	132
#251	ON-1	A36	82	2	9	40	4	137
#252	ON-1	A36	70	3	9	42	4	128
#253	ON-1	A36	84	0	10	41	6	141
#254	ON-1	A36	67	2	9	42	10	125
#255	ON-1	A36	70	2	9	42	5	128
#137	ON-1	A37	84	1	10	41	5	141
#256	ON-1	A37	85	1	10	40	5	141
#258	ON-1	A37	81	1	10	40	5	137
#259	ON-1	A37	83	1	9	41	5	139
#260	ON-1	A37	87	1	9	40	5	142
#261	ON-1	A37	82	1	9	40	5	137
#262	ON-1	A37	84	2	10	41	4	141
#263	ON-1	A37	82	1	10	40	6	139

#264	ON-1	A37	84	1	10	40	5	140
#265	ON-1	A37	84	1	9	40	6	140
#217	QC-1	AM1	97	3	7	39	4	150
#218	QC-1	AM1	97	0	7	42	7	153
#219	QC-1	AM1	99	1	8	43	6	157
#220	QC-1	AM1	97	1	10	39	6	153
#221	QC-1	AM1	101	3	8	39	4	155
#222	QC-1	AM1	97	2	7	40	5	151
#223	QC-1	AM1	99	2	9	41	5	156
#224	QC-1	AM1	98	3	8	40	4	153
#225	QC-1	AM1	99	2	8	39	5	153
#226	QC-1	AM1	96	3	9	40	4	152
#123	ON-1	AM2	72	1	10	44	6	133
#124	ON-1	AM2	71	2	10	42	5	130
#125	ON-1	AM2	73	3	10	41	4	131
#228	ON-1	AM2	72	1	10	41	6	130
#229	ON-1	AM2	72	2	10	44	5	133
#230	ON-1	AM2	73	2	10	43	5	133
#231	ON-1	AM2	72	2	10	41	5	130
#232	ON-1	AM2	72	2	11	42	5	132
#233	ON-1	AM2	71	0	11	41	7	130
#234	ON-1	AM2	72	1	11	41	6	131
#134	ON-1	AM3	82	0	8	40	7	137
#135	ON-1	AM3	69	3	9	44	4	129
#238	ON-1	AM3	84	1	9	40	5	139
#239	ON-1	AM3	69	1	9	42	6	127
#240	ON-1	AM3	69	2	9	44	5	129
#241	ON-1	AM3	69	1	9	42	6	127

#242	ON-1	AM3	84	2	9	40	4	139
#243	ON-1	AM3	72	1	9	43	6	131
#244	ON-1	AM3	82	0	8	40	6	136
#245	ON-1	AM3	83	2	9	40	4	138

Table 5.2: Calculation of the Minimum Number of SNPs required to provide evidence of MSI events within each animal

Animal ID	Age at Sampling (Years)	Minimum Number of SNPs (0.1 SNPs/Genome/Year)[a]	Maximum Number of SNPs (0.5 SNPs/Genome/Year)[b]	Threshold of SNPs required to show evidence of and MSI event (Rounded up from Maximum Number of SNPs calculation)
A19	4	0.4	2.0	2
A20	7.3333	0.7	3.7	4
A21	5.4167	0.5	2.7	3
A22	10.5833	1.1	5.3	6
A23	7.25	0.7	3.6	4
A25	4.3333	0.4	2.2	3
A26	5.5833	0.6	2.8	3
A32	5.25	0.5	2.6	3
A34	7.5	0.8	3.8	4
A36	5.0833	0.5	2.5	3
A37	5.5	0.6	2.8	3
AM1	4.5	0.5	2.3	3
AM2	5.8333	0.6	2.9	3
AM3	5	0.5	2.5	3

[a] *As noted in* [290]
[b] *As noted in* [53]

Table 5.3: Pairwise comparison of SNP differences of isolates within individual animals to provide evidence of microevolution or MSIs.

Animal ID	Isolate Comparison[a]	SNP Difference	MSI Threshold[b]	Evidence of MSI or Microevolution[c]
A19	160 and 168	9	2	MSI
A19	168 and 166	7	2	MSI
A19	166 and 161	6	2	MSI
A19	161 and 165	4	2	MSI
A19	165 and 162	6	2	MSI
A19	162 and 163	2	2	MSI
A19	163 and 167	4	2	MSI
A19	167 and 164	5	2	MSI
A19	164 and 159	1	2	Microevolution
A20	170 and 171	6	4	MSI
A20	171 and 177	2	4	Microevolution
A20	177 and 175	2	4	Microevolution
A20	175 and 176	3	4	Microevolution
A20	176 and 169	0	4	Microevolution
A20	169 and 172	2	4	Microevolution
A20	172 and174	4	4	MSI
A20	174 and 173	2	4	Microevolution
A21	182 and 180	5	3	MSI
A21	180 and 183	8	3	MSI
A21	183 and 188	4	3	MSI
A21	188 and 187	4	3	MSI
A21	187 and 181	1	3	Microevolution
A21	181 and 184	1	3	Microevolution
A21	184 and 186	0	3	Microevolution
A21	186 and 185	0	3	Microevolution
A21	185 and 179	0	3	Microevolution
A22	196 and 190	2	6	Microevolution
A22	190 and 192	4	6	Microevolution
A22	192 and 191	7	6	MSI
A22	191 and 198	2	6	Microevolution
A22	198 and 193	1	6	Microevolution
A22	193 and 195	3	6	Microevolution
A22	195 and 197	7	6	MSI
A22	197 and 194	4	6	Microevolution
A22	194 and 189	3	6	Microevolution
A23	203 and 206	4	4	MSI
A23	206 and 202	4	4	MSI
A23	202 and 208	8	4	MSI

A23	208 and 199	2	4	Microevolution
A23	199 and 204	3	4	Microevolution
A23	204 and 205	4	4	MSI
A23	205 and 201	2	4	Microevolution
A23	201 and 207	7	4	MSI
A23	207 and 200	7	4	MSI
A25	26 and 212	37	3	MSI
A25	212 and 25	6	3	MSI
A25	25 and 28	9	3	MSI
A25	28 and 30	7	3	MSI
A25	30 and 23	0	3	Microevolution
A25	23 and 27	9	3	MSI
A25	27 and 29	7	3	MSI
A25	29 and 24	4	3	MSI
A25	24 and 213	3	3	MSI
A26	33 and 215	2	3	Microevolution
A26	215 and 35	3	3	MSI
A26	35 and 214	5	3	MSI
A26	214 and 38	7	3	MSI
A26	38 and 32	6	3	MSI
A26	32 and 216	6	3	MSI
A26	216 and 36	4	3	MSI
A26	36 and 37	2	3	Microevolution
A26	37 and 34	3	3	MSI
A32	121 and 113	2	3	Microevolution
A32	113 and 116	2	3	Microevolution
A32	116 and 115	2	3	Microevolution
A32	115 and 118	1	3	Microevolution
A32	118 and 119	1	3	Microevolution
A32	119 and 114	2	3	Microevolution
A32	114 and 120	1	3	Microevolution
A32	120 and 122	0	3	Microevolution
A32	122 and 117	0	3	Microevolution
A34	128 and 129	66	4	MSI
A34	129 and 132	3	4	Microevolution
A34	132 and 126	12	4	MSI
A34	126 and 133	4	4	MSI
A34	133 and 127	4	4	MSI
A34	127 and 237	8	4	MSI
A34	237 and 130	3	4	Microevolution
A34	130 and 236	1	4	Microevolution
A34	236 and 131	0	4	Microevolution

A36	253 and 251	2	3	Microevolution
A36	251 and 254	62	3	MSI
A36	254 and 136	3	3	MSI
A36	136 and 252	1	3	Microevolution
A36	252 and 255	0	3	Microevolution
A36	255 and 249	1	3	Microevolution
A36	249 and 248	2	3	Microevolution
A36	248 and 250	3	3	MSI
A36	250 and 247	0	3	Microevolution
A37	260 and 265	8	3	MSI
A37	265 and 256	2	3	Microevolution
A37	256 and 264	1	3	Microevolution
A37	264 and 262	0	3	Microevolution
A37	262 and 263	1	3	Microevolution
A37	263 and 137	2	3	Microevolution
A37	137 and 258	3	3	MSI
A37	258 and 261	1	3	Microevolution
A37	261 and 259	0	3	Microevolution
AM1	219 and 226	5	3	MSI
AM1	226 and 220	3	3	MSI
AM1	220 and 218	2	3	Microevolution
AM1	218 and 217	2	3	Microevolution
AM1	217 and 221	8	3	MSI
AM1	221 and 223	6	3	MSI
AM1	223 and 222	2	3	Microevolution
AM1	222 and 225	2	3	Microevolution
AM1	225 and 224	3	3	MSI
AM2	230 and 125	2	3	Microevolution
AM2	125 and 124	2	3	Microevolution
AM2	124 and 228	1	3	Microevolution
AM2	228 and 233	0	3	Microevolution
AM2	233 and 234	0	3	Microevolution
AM2	234 and 231	0	3	Microevolution
AM2	231 and 232	0	3	Microevolution
AM2	232 and 229	0	3	Microevolution
AM2	229 and 123	0	3	Microevolution
AM3	245 and 238	2	3	Microevolution
AM3	238 and 134	2	3	Microevolution
AM3	134 and 244	4	3	MSI
AM3	244 and 242	1	3	Microevolution
AM3	242 and 241	69	3	MSI
AM3	241 and 135	2	3	Microevolution

AM3	135 and 243	4	3	MSI
AM3	243 and 239	4	3	MSI
AM3	239 and 240	1	3	Microevolution

[a]*Pairwise comparison of isolates as specified based on their relationships in the individual animal phylogenies described in* ***Figure 5.2*** *and* ***Appendix Figure A4****.*
[b]*MSI thresholds as calculated in* ***Table 5.2****.*
[c]*SNP differences greater than the threshold provide evidence of MSI, while SNP differences less than the threshold provide evidence of microevolution.*

Table 5.4: ML-SSR and MIRU-VNTR patterns for 139 isolates from 14 animals as recorded in the MAC-INMV Database

Animal ID	Isolate ID	INMV SSR Type[a]	INMV MIRU Type
A19	#159	MLSSR 10 (6)	INMV 2
A19	#164	MLSSR 10 (6)	INMV 2
A19	#167	MLSSR 10 (6)	INMV 2
A19	#160	MLSSR 10 (7)	INMV 2
A19	#162	MLSSR 10 (7)	INMV 2
A19	#168	MLSSR 10 (7)	INMV 2
A19	#166	MLSSR 7 (2)	INMV 2
A19	#161	MLSSR 7 (3)	INMV 2
A19	#165	MLSSR 93 (2)	INMV 2
A19	#163	UNK	INMV 2
A20	#175	MLSSR 10 (1)	INMV 2
A20	#174	MLSSR 10 (3)	INMV 2
A20	#171	MLSSR 10 (5)	INMV 2
A20	#176	MLSSR 10 (9)	INMV 2
A20	#177	MLSSR 93 (1)	INMV 2
A20	#172	MLSSR 93 (11)	INMV 2
A20	#173	MLSSR 93 (11)	INMV 2
A20	#169	MLSSR 93 (16)	INMV 2
A20	#170	UNK	INMV 2
A21	#179	MLSSR 10 (4)	INMV 2
A21	#180	MLSSR 10 (8)	INMV 2
A21	#183	MLSSR 8 (2)	INMV 2
A21	#184	MLSSR 8 (2)	INMV 2
A21	#187	MLSSR 8 (2)	INMV 2
A21	#185	MLSSR 8 (5)	INMV 2
A21	#182	MLSSR 93 (12)	INMV 2
A21	#186	MLSSR 93 (16)	INMV 2
A21	#188	MLSSR 93 (20)	INMV 2
A21	#181	MLSSR 93 (26)	INMV 2
A22	#198	MLSSR 10 (2)	INMV 2
A22	#191	MLSSR 10 (3)	INMV 2
A22	#194	MLSSR 10 (4)	INMV 2
A22	#190	MLSSR 8 (2)	INMV 2
A22	#196	MLSSR 8 (3)	INMV 2
A22	#192	MLSSR 8 (4)	INMV 2
A22	#197	MLSSR 93 (21)	INMV 2
A22	#193	MLSSR 93 (7)	INMV 2
A22	#195	MLSSR 93 (7)	INMV 2

A22	#189	MLSSR 93 (9)	INMV 2
A23	#202	MLSSR 8 (2)	INMV 2
A23	#206	MLSSR 8 (2)	INMV 2
A23	#208	MLSSR 93 (1)	INMV 2
A23	#199	MLSSR 93 (13)	INMV 2
A23	#204	MLSSR 93 (14)	INMV 2
A23	#200	MLSSR 93 (23)	INMV 2
A23	#201	MLSSR 93 (3)	INMV 2
A23	#205	MLSSR 93 (6)	INMV 2
A23	#203	MLSSR 93 (8)	INMV 2
A23	#207	MLSSR 93 (9)	INMV 2
A25	#23	MLSSR 10 (1)	INMV 2
A25	#24	MLSSR 10 (1)	INMV 2
A25	#27	MLSSR 10 (1)	INMV 2
A25	#28	MLSSR 10 (2)	INMV 2
A25	#213	MLSSR 10 (8)	INMV 2
A25	#30	MLSSR 38	INMV 2
A25	#26	MLSSR 7 (3)	INMV 2
A25	#25	MLSSR 7 (4)	INMV 2
A25	#212	MLSSR 7 (5)	INMV 2
A25	#29	MLSSR 93 (15)	INMV 2
A26	#215	MLSSR 10 (1)	INMV 2
A26	#38	MLSSR 10 (1)	INMV 2
A26	#216	MLSSR 10 (4)	INMV 2
A26	#214	MLSSR 35	INMV 2
A26	#32	MLSSR 38	INMV 2
A26	#35	MLSSR 54	INMV 2
A26	#36	MLSSR 54	INMV 2
A26	#33	MLSSR 7 (1)	INMV 2
A26	#37	MLSSR 8 (2)	INMV 2
A26	#34	MLSSR 93 (4)	INMV 2
A32	#113	N/P 11	INMV 2
A32	#119	N/P 12	INMV 2
A32	#120	N/P 13	INMV 2
A32	#116	N/P 2	INMV 2
A32	#121	N/P 3	INMV 2
A32	#122	N/P 4	INMV 2
A32	#117	N/P 5	INMV 2
A32	#114	N/P 8	INMV 2
A32	#118	N/P 9	INMV 2
A32	#115	UNK	INMV 2
A34	#129	MLSSR 10 (5)	INMV 2

A34	#133	MLSSR 8 (2)	INMV 2
A34	#127	MLSSR 8 (6)	INMV 2
A34	#132	MLSSR 8 (8)	INMV 2
A34	#126	MLSSR 93 (5)	INMV 2
A34	#130	MLSSR 93 (7)	INMV 2
A34	#236	MLSSR 93 (8)	INMV 2
A34	#237	MLSSR 93 (8)	INMV 2
A34	#131	MLSSR 93 (9)	INMV 2
A34	#128	N/P 15	INMV 2
A36	#247	MLSSR 10 (1)	INMV 2
A36	#249	MLSSR 10 (10)	INMV 2
A36	#255	MLSSR 10 (6)	INMV 2
A36	#252	MLSSR 10 (7)	INMV 2
A36	#254	MLSSR 10 (8)	INMV 2
A36	#136	UNK	INMV 2
A36	#248	UNK	INMV 2
A36	#250	UNK	INMV 2
A36	#251	UNK	INMV 2
A36	#253	UNK	INMV 2
A37	#262	N/P 1	INMV 2
A37	#259	N/P 10	INMV 2
A37	#263	N/P 14	INMV 2
A37	#265	N/P 14	INMV 2
A37	#137	N/P 16	INMV 2
A37	#264	N/P 6	INMV 2
A37	#261	N/P 7	INMV 2
A37	#256	UNK	INMV 2
A37	#258	UNK	INMV 2
A37	#260	UNK	INMV 2
AM1	#218	MLSSR 10 (3)	INMV 2
AM1	#225	MLSSR 10 (3)	INMV 2
AM1	#217	MLSSR 10 (6)	INMV 2
AM1	#221	MLSSR 10 (6)	INMV 2
AM1	#219	MLSSR 38	INMV 2
AM1	#220	MLSSR 54	INMV 2
AM1	#224	MLSSR 54	INMV 2
AM1	#222	MLSSR 8 (1)	INMV 2
AM1	#223	MLSSR 93 (1)	INMV 2
AM1	#226	MLSSR 93 (10)	INMV 2
AM2	#231	MLSSR 10 (1)	INMV 2
AM2	#228	MLSSR 8 (5)	INMV 2
AM2	#229	MLSSR 93 (12)	INMV 2

AM2	#125	MLSSR 93 (18)	INMV 2
AM2	#234	MLSSR 93 (19)	INMV 2
AM2	#233	MLSSR 93 (21)	INMV 2
AM2	#124	MLSSR 93 (22)	INMV 2
AM2	#232	MLSSR 93 (24)	INMV 2
AM2	#230	MLSSR 93 (25)	INMV 2
AM2	#123	MLSSR 93 (7)	INMV 2
AM3	#239	MLSSR 10 (7)	INMV 2
AM3	#241	MLSSR 8 (3)	INMV 2
AM3	#240	MLSSR 8 (7)	INMV 2
AM3	#135	MLSSR 93 (16)	INMV 2
AM3	#243	MLSSR 93 (17)	INMV 2
AM3	#134	N/P 17	INMV 2
AM3	#242	N/P 18	INMV 2
AM3	#245	N/P 18	INMV 2
AM3	#244	N/P 19	INMV 2
AM3	#238	UNK	INMV 2

[a]*N/P = New Patterns. Patterns not recognized as a type in the INMV database. Patterns with the same number but different letters would be classified as the same type in the INMV database but have different repeat values. UNK = Unknown. These patterns can't be predicted due to a lack of certainty within specific loci.*

Table 5.5: Isolates examined in the same animal that differ only by SSR1 and SSR2 repeat sizes.

Animal ID	Isolate ID	SSR1 Pattern	SSR1 Reading Frame	SSR2 Pattern	SSR2 Reading Frame
A20	169	17	RF2	12	RF3
A20	176	21	RF3	10	RF1
A21	179	16	RF1	10	RF1
A21	185	18	RF3	11	RF2
A25	23	12	RF3	10	RF1
A25	30	11	RF2	10	RF1
A32	122	15	RF3	12	RF3
A32	117	15	RF3	13	RF1
A36	252	19	RF1	10	RF1
A36	255	18	RF3	10	RF1
A37	262	13	RF1	10	RF1
A37	264	16	RF1	10	RF1
AM2	229	15	RF3	14	RF2
AM2	232	20	RF2	12	RF3

Table 5.6: Classification of homopolymer repeat sizes for SSR1 and SSR2 by reading frame

SSR Locus	Reading Frame	SSR Repeat Size
SSR1	RF-1	4
SSR1	RF-1	7
SSR1	RF-1	10
SSR1	RF-1	13
SSR1	RF-1	16
SSR1	RF-1	19
SSR1	RF-1	22
SSR1	RF-1	25
SSR1	RF-2	5
SSR1	RF-2	8
SSR1	RF-2	11
SSR1	RF-2	14
SSR1	RF-2	17
SSR1	RF-2	20
SSR1	RF-2	23
SSR1	RF-2	26
SSR1	RF-3	6
SSR1	RF-3	9
SSR1	RF-3	12
SSR1	RF-3	15
SSR1	RF-3	18
SSR1	RF-3	21
SSR1	RF-3	24
SSR1	RF-3	27
SSR2	RF-1	7
SSR2	RF-1	10
SSR2	RF-1	13
SSR2	RF-1	16
SSR2	RF-2	8
SSR2	RF-2	11
SSR2	RF-2	14
SSR2	RF-2	17
SSR2	RF-3	9
SSR2	RF-3	12
SSR2	RF-3	15
SSR2	RF-3	18

CHAPTER VI

Conclusions and Future Directions

6.1 Conclusions

In Chapter III of this thesis, a systematic review was conducted to gain a greater understanding of MSIs across the genus *Mycobacterium*, with a specific focus directed to the prevalence of MSI events in both MTBC and NTM infections, whether these infections are observed within both human and animal species, which methods can reliably detect MSIs, and how true MSIs are differentiated from similar events such as re-infections, relapses and microevolution. Results of the systematic review show that MSIs exist in a variety of different mycobacteria, though unsurprisingly, most studies identified MSIs within studies conducted on humans and predominantly focused on *M. tuberculosis* **(Table 3.2)**. In addition, among the variety of methods used to identify MSIs, most studies used a variety of VNTR-based methods for MSI confirmation, able to distinguish between true MSIs and microevolution based on whether multiple alleles can be detected within a certain number of VNTR loci **(Table 3.4)**. Reports from more recent years have shown more reliance on WGS-based analysis, with this change in methodology possibly representing an overall shift as newer technologies are developed and become more widely used for strain typing. Methods based on WGS offer an unprecedented resolution compared to other methods, but lack uniform SNP thresholds, making it difficult to differentiate between true MSI events and microevolution. While the use of specific thresholds may have previously been used to identify MGIs in older methods, such as PFGE or VNTR, a definitive threshold to distinguish strains using WGS methods has not been clearly established. Due to the extensive volume of WGS data and the organism-specific variability that arises due to varying rates of genomic evolution, the establishment of guidelines that work across a spectrum of species becomes difficult [682,683]. Outbreak variability may be affected by both the

species involved, the mutation rates of each strain/species, and the initial genetic distance of any strains of the same species involved in the initial infection event. With the variability of these factors in mind, recent studies recommend that strain-distinguishing thresholds using WGS should be used on a case-by-case basis, with the epidemiological contexts of each outbreak accounted for. This approach was kept in consideration when performing WGS-based analysis in both Chapters IV and V of this thesis. It was also noted in this Chapter that studies on NTM MSIs in humans and animals are limited, but such infections are found across a variety of species. Given the importance of NTM in causing opportunistic and nosocomial infections in humans and diseases in farmed animals, such as JD, the prevalence and impact of MSIs caused by this large and important group of mycobacteria warrant further investigation.

In Chapter IV of this thesis, the genetic variations present in a selection of MAP isolates grown from 67 MAP-shedding cattle from the provinces of QC and ON were analyzed using a selection of typing methods previously identified in Chapter III, including WGS analysis. The assemblies produced using the bioinformatic pipeline were verified as being of high quality without contamination, avoiding the generation of misleading hypotheses due to sequencing or assembly errors. All isolates were confirmed as type II strains, also referred to as cattle-type strains. All strains were subtyped using traditional MLVA methods, such as MIRU-VNTR and ML-SSR. Based on the results provided by the ML-SSR and MIRU-VNTR methods, it was found that the loci used in the ML-SSR method had significantly more discriminatory power than those used by the MIRU-VNTR method, as they were able to identify more unique patterns both across the entire dataset and within each of the individual herds. However, the limited discriminatory capacity of both MLVA methods did not allow for tracking of inter-herd MAP transmission. The core SNP-based analysis was the only approach leading to the assignment of

individual signatures to each isolate and allowed for the documentation of disease transmission across herds. This was further supported by animal movement records, which noted when cattle were introduced to specific herds and if they had been delivered from another farm. The structure of the 67 isolates examined in this Chapter did not correlate with animal phenotypic data, though the presence of genetic variations in several virulence factor genes, namely from the *PE/PPE*, *mce*, and *mmpL* families, may potentially explain differential antigenic or pathogenetic responses.

In Chapter V of this thesis, an in-depth investigation of the presence of MAP MGIs in dairy cattle was conducted, selecting up to 10 isolates from 14 cattle, and analyzing them using a WGS-based approach. By analyzing the genome sequences of multiple MAP isolates from the same animal, the identification of MGIs and the ability to differentiate between true MSIs and microevolution-based events is possible. MGIs were able to be classified into different categories based on the genetic relatedness of MAP isolates, ranging from those caused by MAP strains with result in a diverse SNP-based phylogeny, to instances where the isolates were much more closely related. It was noted in both this Chapter and Chapter IV, that the ML-SSR method allowed for enhanced discrimination between MAP isolates when complemented with SNP analysis, and both methods were more discriminatory than MIRU-VNTR. Maximum diversity in both Chapters was noted within ML-SSR loci SSR1 and SSR2, and as in some cases in Chapter V, were the only sequences that differed between otherwise isogenic MAP isolates **(Figure 5.2** and **Table 5.5**). It was noted that all *ORF1* sequences from other MAC subspecies corresponded to RF-1, almost all of which contained the seven-nucleotide "GGCGGGG" sequence instead of the "G" homopolymer present in MAP. Analysis of the SSR1 and SSR2 repeat sizes within all available MAP strains confirmed that the most common SSR1 repeat size in MAP was seven

(Figures 5.3, **5.4** and **Appendix Table A1)**, which resulted in a protein sequence that matches the predicted *ORF1* sequence [67,633,634]. These results also suggest the possibility that a C→G transversion event in an ancestral MAC bacteria resulted in the establishment of the SSR1 mononucleotide homopolymer observed in MAP *ORF1* [668,669]. This, in turn, resulted in the variety of repeat sizes reported for SSR1 in MAP, while also affecting the reading frame of *ORF1* **(Appendix Table A12)**. In the case of *ORF2*, the *M. avium* subsp. *avium*, *M. avium* subsp. *hominissuis* and *M. avium* subsp. *lepraemurium* orthologs also contain the SSR2 homopolymer, where the *M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis* sequences exhibit variations similar to those observed in MAP. As with MAP, this suggests that *M. avium* subsp. *avium* and *M. avium* subsp. *hominissuis* may have a similar capacity to produce different gene products using alternate reading frames.

Variations caused in the amino acid sequences and corresponding structures of MAP proteins due to the presence of variable SSRs in cognate genes have not received much attention. The ORF1 and ORF2 variants predicted in Chapter V reinforce the importance of variable genetic elements in affecting protein products, beyond use as molecular targets for strain typing methods. While not much is known about the functions of ORF1 and ORF2, it can be hypothesized that changes in the cognate ORF1 product caused by SSR-associated frame shifting could potentially affect the interaction of MAP with the host. It remains unknown how these SSRs change or evolve in infecting MAP isolates, but for the purposes of this work, MAP isolates with different SSRs are currently classified as MGIs for now, as they are technically not isogenic.

6.2 Future Directions

As mentioned in Chapter III, there is an underrepresentation of studies examining MSIs throughout the genus *Mycobacterium* in both NTMs and infections involving non-human hosts. Both groups are the focus of this work, though examination of MSIs in non-MAP NTMs and animal hosts beyond cattle are also of interest. As noted in Chapter IV, variation in the *PE/PPE*, *mce*, and *mmpL* virulence factor gene families may have an impact on antigenic variation and pathogenetic responses of MAP within some cattle. However, work involving the direct examination of such mutations, and the impact such changes have on the immune response of cattle, must be done to determine the validity of such a claim. The use of bovine monocyte-derived macrophage, in combination with strains that contain mutations of interest within these virulence factor families, may allow for the determination of the exact impact of these mutations. Work conducted in Chapter V showed variation within both SSR1 and SSR2 and the impact that alteration of this homopolymer sequence has on ORF1 and ORF2. Of particular interest were select instances where strains differed only at one or both of these loci, with no other differences identified in either strain **(Table 5.5)**. Further work using these almost identical strains will allow for the direct examination of how these repeat differences, and subsequence reading frame alterations, impact infection.

In addition to the future work described above, current work not ready for inclusion in this thesis may also offer new insights into the impact of MAP MSIs and how they change over time. In some ongoing work, a selection of ten isolates taken at two time points from seven cattle are undergoing the same bioinformatic analysis described in Chapters IV and V of this thesis. This analysis, and further examination using molecular clock analysis, may help identify how both individual strains and MSI events, are change between two points in time. A long-term goal

of this project is the comparison between both MAP and host animal genomes to identify correlations between the two for factors such as shedding value, the speed and severity of disease progression, and variant identification in the hope of identifying host factors that promote MAP infection resistance or MAP targets which may allow for the development of a viable vaccine.

BIBLIOGRAPHY

1. Government of Canada, A. and A.-F. Canada's dairy industry at a glance. https://agriculture.canada.ca/en/sector/animal-industry/canadian-dairy-information-centre/dairy-industry (2022).
2. Aghamohammadi, M. *et al.* Herd-Level Mastitis-Associated Costs on Canadian Dairy Farms. *Frontiers in Veterinary Science* **5**, (2018).
3. Gohary, K. *et al.* The cost of a case of subclinical ketosis in Canadian dairy herds. *Can Vet J* **57**, 728–732 (2016).
4. Roche, S. M. *et al.* Cost-benefit of implementing a participatory extension model for improving on-farm adoption of Johne's disease control recommendations. *J Dairy Sci* **103**, 451–472 (2020).
5. Abubakar, I., Myhill, D., Aliyu, S. H. & Hunter, P. R. Detection of *Mycobacterium avium* subspecies *paratuberculosis* from patients with Crohn's disease using nucleic acid-based techniques: a systematic review and meta-analysis. *Inflamm Bowel Dis* **14**, 401–410 (2008).
6. Singh, S. & Gopinath, K. *Mycobacterium* avium subspecies *paratuberculosis* and Crohn's Regional Ileitis: How Strong is Association? *J Lab Physicians* **3**, 69–74 (2011).
7. Mintz, M. J. & Lukin, D. J. *Mycobacterium avium* subspecies *paratuberculosis* (MAP) and Crohn's disease: the debate continues. *Transl Gastroenterol Hepatol* **8**, 28 (2023).

8. Atreya, R. *et al.* Facts, myths and hypotheses on the zoonotic nature of *Mycobacterium avium* subspecies *paratuberculosis*. *International Journal of Medical Microbiology* **304**, 858–867 (2014).
9. Rasmussen, P., Barkema, H. W., Mason, S., Beaulieu, E. & Hall, D. C. Economic losses due to Johne's disease (paratuberculosis) in dairy cattle. *Journal of Dairy Science* **104**, 3123–3143 (2021).
10. Ahmed, I., Jabeen, K. & Hasan, R. Identification of non-tuberculous mycobacteria isolated from clinical specimens at a tertiary care hospital: a cross-sectional study. *BMC Infect Dis* **13**, 493 (2013).
11. Parte, A. C. LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int J Syst Evol Microbiol* **68**, 1825–1829 (2018).
12. Primm, T. P., Lucero, C. A. & Falkinham, J. O. Health Impacts of Environmental Mycobacteria. *Clin Microbiol Rev* **17**, 98–106 (2004).
13. Bespiatykh, D., Bespyatykh, J., Mokrousov, I. & Shitikov, E. A Comprehensive Map of *Mycobacterium tuberculosis* Complex Regions of Difference. *mSphere* **6**, e0053521 (2021).
14. Singh, P. *et al.* Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. *Proc Natl Acad Sci U S A* **112**, 4459–4464 (2015).
15. Siddiqi, S. H. Atypical mycobacteria. *J Pak Med Assoc* **28**, 102–110 (1978).
16. Bachmann, N. L. *et al.* Key Transitions in the Evolution of Rapid and Slow Growing Mycobacteria Identified by Comparative Genomics. *Frontiers in Microbiology* **10**, (2020).

17. Kim, C.-J. *et al.* Differentiating rapid- and slow-growing mycobacteria by difference in time to growth detection in liquid media. *Diagnostic Microbiology and Infectious Disease* **75**, 73–76 (2013).

18. Runyon, E. H. Typical Mycobacteria: Their classification. *Am Rev Respir Dis* **91**, 288–289 (1965).

19. Gupta, R. S., Lo, B. & Son, J. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Frontiers in Microbiology* **9**, (2018).

20. Nouioui, I. *et al.* Genome-Based Taxonomic Classification of the Phylum Actinobacteria. *Frontiers in Microbiology* **9**, (2018).

21. Falkinham, J. O. Environmental Sources of Nontuberculous Mycobacteria. *Clinics in Chest Medicine* **36**, 35–41 (2015).

22. Jeon, D. Infection Source and Epidemiology of Nontuberculous Mycobacterial Lung Disease. *Tuberc Respir Dis (Seoul)* **82**, 94–101 (2019).

23. Falkinham, J. O. *Mycobacterium avium* Complex (MAC) in Water Distribution Systems and Household Plumbing in the United States. *Water* **12**, 3338 (2020).

24. Katila, M. L. *et al.* Isolation of potentially pathogenic mycobacteria in the Finnish environment. *Scand J Infect Dis Suppl* **98**, 9–11 (1995).

25. Parker, B. C., Ford, M. A., Gruft, H. & Falkinham, J. O. Epidemiology of infection by nontuberculous mycobacteria. IV. Preferential aerosolization of *Mycobacterium intracellulare* from natural waters. *Am Rev Respir Dis* **128**, 652–656 (1983).

26. Tzou, C. L. *et al.* Association between *Mycobacterium avium* Complex Pulmonary Disease and Mycobacteria in Home Water and Soil. *Ann Am Thorac Soc* **17**, 57–62 (2020).

27. Walsh, C. M., Gebert, M. J., Delgado-Baquerizo, M., Maestre, F. T. & Fierer, N. A Global Survey of Mycobacterial Diversity in Soil. *Applied and Environmental Microbiology* **85**, e01180-19 (2019).

28. Yajko, D. M. *et al. Mycobacterium avium* complex in water, food, and soil samples collected from the environment of HIV-infected individuals. *J Acquir Immune Defic Syndr Hum Retrovirol* **9**, 176–182 (1995).

29. Falkinham, J. O. Mycobacterial Aerosols and Respiratory Disease. *Emerg Infect Dis* **9**, 763–767 (2003).

30. Ichiyama, S., Shimokata, K. & Tsukamura, M. The isolation of *Mycobacterium avium* complex from soil, water, and dusts. *Microbiol Immunol* **32**, 733–739 (1988).

31. Diel, R., Lipman, M. & Hoefsloot, W. High mortality in patients with *Mycobacterium avium* complex lung disease: a systematic review. *BMC Infectious Diseases* **18**, 206 (2018).

32. Field, S. K., Fisher, D. & Cowie, R. L. *Mycobacterium avium* complex pulmonary disease in patients without HIV infection. *Chest* **126**, 566–581 (2004).

33. Bruijnesteijn van Coppenraet, L. E. S., de Haas, P. E. W., Lindeboom, J. A., Kuijper, E. J. & van Soolingen, D. Lymphadenitis in children is caused by *Mycobacterium avium hominissuis* and not related to 'bird tuberculosis'. *Eur J Clin Microbiol Infect Dis* **27**, 293–299 (2008).

34. Christensen, J. B. & Koeppe, J. *Mycobacterium avium* Complex Cervical Lymphadenitis in an Immunocompetent Adult. *Clin Vaccine Immunol* **17**, 1488–1490 (2010).

35. Thegerström, J. *et al. Mycobacterium avium* Lymphadenopathy among Children, Sweden. *Emerg Infect Dis* **14**, 661–663 (2008).

36. Ayoade, F., Cotelingam, J. & Joel Chandranesan, A. S. Disseminated *Mycobacterium avium-intracellulare* Complex Infection Presenting With Disseminated Intravascular Coagulation in an AIDS Patient. *J Investig Med High Impact Case Rep* **5**, 2324709617740904 (2017).

37. Horsburgh, C. R., Jr. The Pathophysiology of Disseminated *Mycobacterium avium* Complex Disease in AIDS. *The Journal of Infectious Diseases* **179**, S461–S465 (1999).

38. Ito, Y. *et al.* Disseminated *Mycobacterium avium* Complex Infection in a Patient Treated With Immunosuppressants. *ACG Case Reports Journal* **10**, e01033 (2023).

39. van Ingen, J., Turenne, C. Y., Tortoli, E., Wallace, R. J. & Brown-Elliott, B. A. A definition of the *Mycobacterium avium* complex for taxonomical and clinical purposes, a review. *Int J Syst Evol Microbiol* **68**, 3666–3677 (2018).

40. Abe, Y. *et al. Mycobacterium senriense* sp. nov., a slowly growing, non-scotochromogenic species, isolated from sputum of an elderly man. *International Journal of Systematic and Evolutionary Microbiology* **72**, 005378 (2022).

41. Bang, D. *et al. Mycobacterium arosiense* sp. nov., a slowly growing, scotochromogenic species causing osteomyelitis in an immunocompromised child.

International Journal of Systematic and Evolutionary Microbiology **58**, 2398–2402 (2008).

42. Ben Salah, I., Cayrou, C., Raoult, D. & Drancourt, M. *Mycobacterium marseillense* sp. nov., *Mycobacterium timonense* sp. nov. and *Mycobacterium bouchedurhonense* sp. nov., members of the *Mycobacterium avium* complex. *Int J Syst Evol Microbiol* **59**, 2803–2808 (2009).
43. Kim, B.-J. *et al. Mycobacterium yongonense* sp. nov., a slow-growing non-chromogenic species closely related to *Mycobacterium intracellulare*. *Int J Syst Evol Microbiol* **63**, 192–199 (2013).
44. Lee, S.-Y. *et al. Mycobacterium paraintracellulare* sp. nov., for the genotype INT-1 of *Mycobacterium intracellulare*. *International Journal of Systematic and Evolutionary Microbiology* **66**, 3132–3141 (2016).
45. Murcia, M. I., Tortoli, E., Menendez, M. C., Palenque, E. & Garcia, M. J. *Mycobacterium colombiense* sp. nov., a novel member of the *Mycobacterium avium* complex and description of MAC-X as a new ITS genetic variant. *Int J Syst Evol Microbiol* **56**, 2049–2054 (2006).
46. Tortoli, E. *et al.* Proposal to elevate the genetic variant MAC-A, included in the *Mycobacterium avium* complex, to species rank as *Mycobacterium chimaera* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **54**, 1277–1285 (2004).
47. van Ingen, J. *et al.* Proposal to elevate *Mycobacterium avium* complex ITS sequevar MAC-Q to *Mycobacterium vulneris* sp. nov. *Int J Syst Evol Microbiol* **59**, 2277–2282 (2009).

48. van Ingen, J. *et al. Mycobacterium mantenii* sp. nov., a pathogenic, slowly growing, scotochromogenic species. *Int J Syst Evol Microbiol* **59**, 2782–2787 (2009).

49. Toney, N. C., Toney, S. R. & Butler, W. R. Utility of high-performance liquid chromatography analysis of mycolic acids and partial 16S rRNA gene sequencing for routine identification of *Mycobacterium* spp. in a national reference laboratory. *Diagnostic Microbiology and Infectious Disease* **67**, 143–152 (2010).

50. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology* **57**, 81–91 (2007).

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

52. Tortoli, E. *et al.* Genome-based taxonomic revision detects a number of synonymous taxa in the genus *Mycobacterium*. *Infection, Genetics and Evolution* **75**, 103983 (2019).

53. Bryant, J. M. *et al.* Phylogenomic exploration of the relationships between strains of *Mycobacterium avium* subspecies *paratuberculosis*. *BMC Genomics* **17**, 79 (2016).

54. Radomski, N. *et al.* Determination of Genotypic Diversity of *Mycobacterium avium* Subspecies from Human and Animal Origins by Mycobacterial Interspersed Repetitive-Unit-Variable-Number Tandem-Repeat and IS1311 Restriction Fragment Length Polymorphism Typing Methods. *J Clin Microbiol* **48**, 1026–1034 (2010).

55. Johansen, T. B., Djønne, B., Jensen, M. R. & Olsen, I. Distribution of *IS1311* and *IS1245* in *Mycobacterium avium* Subspecies Revisited. *Journal of Clinical Microbiology* **43**, 2500–2502 (2005).

56. Semret, M. *et al.* Genomic Polymorphisms for *Mycobacterium avium* subsp. *paratuberculosis* Diagnostics. *Journal of Clinical Microbiology* **43**, 3704–3712 (2005).

57. Bannantine, J. P., Conde, C., Bayles, D. O., Branger, M. & Biet, F. Genetic Diversity Among *Mycobacterium avium* Subspecies Revealed by Analysis of Complete Genome Sequences. *Front Microbiol* **11**, 1701 (2020).

58. Dhama, K. *et al.* Tuberculosis in Birds: Insights into the *Mycobacterium avium* Infections. *Vet Med Int* **2011**, 712369 (2011).

59. Tell, L. A., Woods, L. & Cromie, R. L. Mycobacteriosis in birds. *Rev Sci Tech* **20**, 180–203 (2001).

60. Zhu, L. *et al.* Isolation, Identification, and Characterization of a New Highly Pathogenic Field Isolate of *Mycobacterium avium* spp. *avium*. *Frontiers in Veterinary Science* **4**, (2018).

61. Agdestein, A. *et al.* A comparative study of *Mycobacterium avium* subsp. *avium* and *Mycobacterium avium* subsp. *hominissuis* in experimentally infected pigs. *BMC Veterinary Research* **8**, 11 (2012).

62. Moreno, B., Aduriz, G., Garrido, J. M., Sevilla, I. & Juste, R. A. Disseminated *Mycobacterium avium* subsp. *avium* infection in a pet Korean squirrel (Sciuris vulgaris coreae). *Vet Pathol* **44**, 123–125 (2007).

63. Slany, M., Ulmann, V. & Slana, I. Avian Mycobacteriosis: Still Existing Threat to Humans. *BioMed Research International* **2016**, e4387461 (2016).

64. Rónai, Z., Csivincsik, Á. & Dán, Á. Molecular Identification of *Mycobacterium avium* subsp. *silvaticum* by Duplex High-Resolution Melt Analysis and Subspecies-Specific Real-Time PCR. *J Clin Microbiol* **53**, 1582–1587 (2015).

65. Saxegaard, F. & Baess, I. Relationship between *Mycobacterium avium*, *Mycobacterium paratuberculosis* and "wood pigeon mycobacteria". *APMIS* **96**, 37–42 (1988).

66. Thorel, M. F., Krichevsky, M. & Lévy-Frébault, V. V. Numerical taxonomy of mycobactin-dependent mycobacteria, emended description of *Mycobacterium avium*, and description of *Mycobacterium avium* subsp. *avium* subsp. nov., *Mycobacterium avium* subsp. *paratuberculosis* subsp. nov., and *Mycobacterium avium* subsp. *silvaticum* subsp. nov. *Int J Syst Bacteriol* **40**, 254–260 (1990).

67. Turenne, C. Y., Collins, D. M., Alexander, D. C. & Behr, M. A. *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* subsp. *avium* Are Independently Evolved Pathogenic Clones of a Much Broader Group of M. avium Organisms. *J Bacteriol* **190**, 2479–2487 (2008).

68. Rindi, L. & Garzelli, C. Genetic diversity and phylogeny of *Mycobacterium avium*. *Infection, Genetics and Evolution* **21**, 375–383 (2014).

69. Haist, V. *et al. Mycobacterium avium* subsp. *hominissuis* Infection in 2 Pet Dogs, Germany. *Emerg Infect Dis* **14**, 988–990 (2008).

70. Chaisson, R. E., Moore, R. D., Richman, D. D., Keruly, J. & Creagh, T. Incidence and natural history of *Mycobacterium avium*-complex infections in patients with

advanced human immunodeficiency virus disease treated with zidovudine. The Zidovudine Epidemiology Study Group. *Am Rev Respir Dis* **146**, 285–289 (1992).

71. Dragset, M. S. *et al.* Global Assessment of *Mycobacterium avium* subsp. *hominissuis* Genetic Requirement for Growth and Virulence. *mSystems* **4**, e00402-19 (2019).
72. Falkinham, J. O. Epidemiology of infection by nontuberculous mycobacteria. *Clin Microbiol Rev* **9**, 177–215 (1996).
73. Becerril-Villanueva, E. *et al.* Chronic infection with *Mycobacterium lepraemurium* induces alterations in the hippocampus associated with memory loss. *Sci Rep* **8**, 9063 (2018).
74. O'Brien, C. R. *et al.* Feline leprosy due to *Mycobacterium lepraemurium*: Further clinical and molecular characterisation of 23 previously reported cases and an additional 42 cases. *Journal of Feline Medicine and Surgery* **19**, 737–746 (2017).
75. Rojas-Espinosa, O. & Løvik, M. *Mycobacterium leprae* and *Mycobacterium lepraemurium* infections in domestic and wild animals. *Rev Sci Tech* **20**, 219–251 (2001).
76. Stefansky, W. Zabolevanija u krys, vyzvannyja kislotoupornoj palotsjkoj. *Rus Vratsj* **47**, 1726–1727 (1902).
77. Benjak, A. *et al.* Insights from the Genome Sequence of *Mycobacterium lepraemurium*: Massive Gene Decay and Reductive Evolution. *mBio* **8**, e01283-17 (2017).
78. Singh, A. V., Singh, S. V., Singh, P. K. & Sohal, J. S. Genotype diversity in Indian isolates of *Mycobacterium avium* subspecies *paratuberculosis* recovered from

domestic and wild ruminants from different agro-climatic regions. *Comparative Immunology, Microbiology and Infectious Diseases* **33**, e127–e131 (2010).

79. Robino, P. *et al.* Identification of *Mycobacterium avium* subsp. *paratuberculosis* in wild cervids (*Cervus elaphus hippelaphus* and *Capreolus capreolus*) from Northwestern Italy. *Eur J Wildl Res* **54**, 357–360 (2008).
80. Roller, M. *et al. Mycobacterium avium* subspecies *paratuberculosis* Infection in Zoo Animals: A Review of Susceptibility and Disease Process. *Frontiers in Veterinary Science* **7**, (2020).
81. Honap, S. *et al.* Anti-*Mycobacterium paratuberculosis* (MAP) therapy for Crohn's disease: an overview and update. *Frontline Gastroenterology* **12**, 397–403 (2021).
82. McNees, A. L., Markesich, D., Zayyani, N. R. & Graham, D. Y. *Mycobacterium paratuberculosis* as a cause of Crohn's disease. *Expert Rev Gastroenterol Hepatol* **9**, 1523–1534 (2015).
83. Pierce, E. S. Ulcerative colitis and Crohn's disease: is *Mycobacterium avium* subspecies *paratuberculosis* the common villain? *Gut Pathogens* **2**, 21 (2010).
84. Sechi, L. A. & Dow, C. T. *Mycobacterium avium* ss. *paratuberculosis* Zoonosis – The Hundred Year War – Beyond Crohn's Disease. *Frontiers in Immunology* **6**, (2015).
85. Dow, C. T. & Ellingson, J. L. E. Detection of *Mycobacterium avium ss. paratuberculosis* in Blau Syndrome Tissues. *Autoimmune Diseases* **2010**, e127692 (2010).

86. Cossu, A. *et al.* MAP3738c and MptD are specific tags of Mycobacterium avium subsp. paratuberculosis infection in type I diabetes mellitus. *Clinical Immunology* **141**, 49–57 (2011).

87. Dow, C. T. Paratuberculosis and Type I diabetes Is this the trigger? *Medical Hypotheses* **67**, 782–785 (2006).

88. Masala, S. *et al.* Antibodies Recognizing *Mycobacterium avium paratuberculosis* Epitopes Cross-React with the Beta-Cell Antigen ZnT8 in Sardinian Type 1 Diabetic Patients. *PLOS ONE* **6**, e26931 (2011).

89. Rosu, V. *et al.* Specific Immunoassays Confirm Association of *Mycobacterium avium* subsp. *paratuberculosis* with Type-1 but Not Type-2 Diabetes Mellitus. *PLOS ONE* **4**, e4386 (2009).

90. Sechi, L. A. *et al. Mycobacterium avium* subspecies *paratuberculosis* Bacteremia in Type 1 Diabetes Mellitus: An Infectious Trigger? *Clinical Infectious Diseases* **46**, 148–149 (2008).

91. Sechi, L. A. *et al.* Humoral Immune Responses of Type 1 Diabetes Patients to *Mycobacterium avium* subsp. *paratuberculosis* Lend Support to the Infectious Trigger Hypothesis. *Clinical and Vaccine Immunology* **15**, 320–326 (2008).

92. Masala, S., Cossu, D., Palermo, M. & Sechi, L. A. Recognition of Zinc Transporter 8 and MAP3865c Homologous Epitopes by Hashimoto's Thyroiditis Subjects from Sardinia: A Common Target with Type 1 Diabetes? *PLOS ONE* **9**, e97621 (2014).

93. Moghadam, M., Ghaemi, E. A., Akbari, H., Razavi Nikoo, H. & Zamani, S. *Mycobacterium avium* subsp. *paratuberculosis* and Hashimoto's thyroiditis: Is MAP the trigger? *Front Cell Infect Microbiol* **12**, 972929 (2022).

94. Sisto, M. *et al.* Proposing a relationship between *Mycobacterium avium* subspecies *paratuberculosis* infection and Hashimoto's thyroiditis. *Scandinavian Journal of Infectious Diseases* **42**, 787–790 (2010).

95. Cossu, D. *et al.* Association of *Mycobacterium avium* subsp. *paratuberculosis* with Multiple Sclerosis in Sardinian Patients. *PLOS ONE* **6**, e18482 (2011).

96. Cossu, D. *et al.* Humoral response against host-mimetic homologous epitopes of *Mycobacterium avium* subsp. *paratuberculosis* in Japanese multiple sclerosis patients. *Sci Rep* **6**, 29227 (2016).

97. Cossu, D. *et al.* Evaluation of the humoral response against mycobacterial peptides, homologous to MOG35–55, in multiple sclerosis patients. *Journal of the Neurological Sciences* **347**, 78–81 (2014).

98. Cossu, D. *et al.* Antigenic epitopes of MAP2694 homologous to T-cell receptor gamma-chain are highly recognized in multiple sclerosis Sardinian patients. *Molecular Immunology* **57**, 138–140 (2014).

99. Davide, C., Yokoyama, K., Nobutaka, H. & A Sechi, L. From Sardinia to Japan: update on the role of MAP in multiple sclerosis. *Future Microbiology* **14**, 643–646 (2019).

100. Frau, J. *et al.* Role of interferon-beta in *Mycobacterium avium* subspecies *paratuberculosis* antibody response in Sardinian MS patients. *Journal of the Neurological Sciences* **349**, 249–250 (2015).

101. Mameli, G. *et al.* EBNA-1 IgG titers in Sardinian multiple sclerosis patients and controls. *Journal of Neuroimmunology* **264**, 120–122 (2013).

102. Otsubo, S. *et al.* Seroprevalence of IgG1 and IgG4 Class Antibodies Against *Mycobacterium avium* subsp. *paratuberculosis* in Japanese Population. *Foodborne Pathogens and Disease* **12**, 851–856 (2015).

103. Rogall, T., Wolters, J., Flohr, T. & Böttger, E. C. Towards a Phylogeny and Definition of Species at the Molecular Level within the Genus *Mycobacterium*. *International Journal of Systematic and Evolutionary Microbiology* **40**, 323–330 (1990).

104. Stevenson, K. Genetic diversity of *Mycobacterium avium* subspecies *paratuberculosis* and the influence of strain type on infection and pathogenesis: a review. *Veterinary Research* **46**, 64 (2015).

105. Dane, H., Stewart, L. D. & Grant, I. R. Culture of *Mycobacterium avium* subsp. *paratuberculosis*: challenges, limitations and future prospects. *Journal of Applied Microbiology* **134**, lxac017 (2023).

106. Francis, J., Macturk, H. M., Madinaveitia, J. & Snow, G. A. Mycobactin, a growth factor for *Mycobacterium johnei*. 1. Isolation from *Mycobacterium phlei*. *Biochem J* **55**, 596–607 (1953).

107. Rathnaiah, G. *et al.* Pathogenesis, Molecular Genetics, and Genomics of *Mycobacterium avium* subsp. *paratuberculosis*, the Etiologic Agent of Johne's Disease. *Front Vet Sci* **4**, 187 (2017).

108. McMahon, M. D., Rush, J. S. & Thomas, M. G. Analyses of *MbtB*, *MbtE*, and *MbtF* Suggest Revisions to the Mycobactin Biosynthesis Pathway in *Mycobacterium tuberculosis*. *J Bacteriol* **194**, 2809–2818 (2012).

109. Reddy, P. V. *et al.* Disruption of Mycobactin Biosynthesis Leads to Attenuation of *Mycobacterium tuberculosis* for Growth and Virulence. *The Journal of Infectious Diseases* **208**, 1255–1265 (2013).

110. Snow, G. A. Mycobactins: iron-chelating growth factors from mycobacteria. *Bacteriol Rev* **34**, 99–125 (1970).

111. Kruijf, M. de, Coffey, A. & O'Mahony, J. The investigation of the truncated *mbtA* gene within the mycobactin cluster of *Mycobacterium avium* subspecies *paratuberculosis* as a novel diagnostic marker for real-time PCR. *Journal of Microbiological Methods* **136**, 40–48 (2017).

112. Li, L. *et al.* The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc Natl Acad Sci U S A* **102**, 12344–12349 (2005).

113. Shoyama, F. M., Janetanakit, T., Bannantine, J. P., Barletta, R. G. & Sreevatsan, S. Elucidating the Regulon of a Fur-like Protein in *Mycobacterium avium* subsp. *paratuberculosis* (MAP). *Frontiers in Microbiology* **11**, (2020).

114. Wang, J. *et al.* Iron Acquisition in *Mycobacterium avium* subsp. *paratuberculosis*. *J Bacteriol* **198**, 857–866 (2016).

115. Green, E. P. *et al.* Sequence and characteristics of *IS900*, an insertion element identified in a human Crohn's disease isolate of *Mycobacterium paratuberculosis*. *Nucleic Acids Res* **17**, 9063–9073 (1989).

116. Conde, C. *et al.* Whole-Genome Analysis of *Mycobacterium avium* subsp. *paratuberculosis IS900* Insertions Reveals Strain Type-Specific Modalities. *Front Microbiol* **12**, 660002 (2021).

117. Paustian, M. L., Amonsin, A., Kapur, V. & Bannantine, J. P. Characterization of Novel Coding Sequences Specific to *Mycobacterium avium* subsp. *paratuberculosis*: Implications for Diagnosis of Johne's Disease. *Journal of Clinical Microbiology* **42**, 2675–2681 (2004).

118. Rani, M. *et al. ISMap02* element targeted nested polymerase chain in the detection of *Mycobacterium avium* subsp. *paratuberculosis* in fecal samples of cattle and buffaloes. *Vet World* **11**, 397–401 (2018).

119. Cousins, D. V., Williams, S. N., Hope, A. & Eamens, G. J. DNA fingerprinting of Australian isolates of *Mycobacterium avium* subsp. *paratuberculosis* using *IS900* RFLP. *Aust Vet J* **78**, 184–190 (2000).

120. Cousins, D. V. *et al.* Mycobacteria distinct from *Mycobacterium avium* subsp. *paratuberculosis* isolated from the faeces of ruminants possess *IS900*-like sequences detectable *IS900* polymerase chain reaction: implications for diagnosis. *Mol Cell Probes* **13**, 431–442 (1999).

121. Englund, S., Bölske, G. & Johansson, K.-E. An *IS900*-like sequence found in a *Mycobacterium* sp. other than *Mycobacterium avium* subsp. *paratuberculosis*. *FEMS Microbiology Letters* **209**, 267–271 (2002).

122. Bannantine, J. P., Stabel, J. R., Bayles, D. O. & Biet, F. Improved DNA Amplification of the Hallmark *IS900* Element in *Mycobacterium avium* subsp. *paratuberculosis*: a Reexamination Based on Whole-Genome Sequence Analysis. *Appl Environ Microbiol* **89**, e01682-22 (2023).

123. Vary, P. H., Andersen, P. R., Green, E., Hermon-Taylor, J. & McFadden, J. J. Use of highly specific DNA probes and the polymerase chain reaction to detect

Mycobacterium paratuberculosis in Johne's disease. *J Clin Microbiol* **28**, 933–937 (1990).

124. Sohal, J. S. *et al. Mycobacterium avium* subspecies *paratuberculosis* diagnosis and geno-typing: Genomic insights. *Microbiological Research* **164**, 330–337 (2009).
125. Stabel, J. R. & Bannantine, J. P. Development of a Nested PCR Method Targeting a Unique Multicopy Element, *ISMap02*, for Detection of *Mycobacterium avium* subsp. *paratuberculosis* in Fecal Samples. *Journal of Clinical Microbiology* **43**, 4744–4750 (2005).
126. Park, H.-T., Park, H.-E., Jung, Y.-H. & Yoo, H. S. An *ISMap02*-like insertion sequence in *Mycobacterium* spp. interferes with specific detection of *Mycobacterium avium* subsp. *paratuberculosis*. *Veterinary Microbiology* **216**, 1–6 (2018).
127. Bannantine, J. P., Stabel, J. R., Bayles, D. O., Conde, C. & Biet, F. Diagnostic Sequences That Distinguish *M. avium* Subspecies Strains. *Front Vet Sci* **7**, 620094 (2021).
128. Bannantine, J. P. & Stabel, J. R. *HspX* is present within *Mycobacterium paratuberculosis*-infected macrophages and is recognized by sera from some infected cattle. *Vet Microbiol* **76**, 343–358 (2000).
129. Ellingson, J. L., Bolin, C. A. & Stabel, J. R. Identification of a gene unique to *Mycobacterium avium* subspecies *paratuberculosis* and application to diagnosis of paratuberculosis. *Mol Cell Probes* **12**, 133–142 (1998).
130. Thirumalapura, N. R., Feria, W., Hue, E., Zellers, C. & Tewari, D. Evaluation of a high-throughput nucleic acid extraction method for the detection of *Mycobacterium*

avium subsp. *paratuberculosis* in bovine fecal samples by PCR. *J Vet Diagn Invest* **33**, 375–378 (2021).

131. Sidoti, F. *et al.* Validation and standardization of *IS900* and *F57* real-time quantitative PCR assays for the specific detection and quantification of *Mycobacterium avium* subsp. *paratuberculosis*. *Can. J. Microbiol.* **57**, 347–354 (2011).
132. Slana, I., Kralik, P., Kralova, A. & Pavlik, I. On-farm spread of *Mycobacterium avium* subsp. *paratuberculosis* in raw milk studied by *IS900* and *F57* competitive real time quantitative PCR and culture examination. *International Journal of Food Microbiology* **128**, 250–257 (2008).
133. Tasara, T. & Stephan, R. Development of an *F57* Sequence-Based Real-Time PCR Assay for Detection of *Mycobacterium avium* subsp. *paratuberculosis* in Milk. *Applied and Environmental Microbiology* **71**, 5957–5968 (2005).
134. Bannantine, J. P., Baechler, E., Zhang, Q., Li, L. & Kapur, V. Genome Scale Comparison of *Mycobacterium avium* subsp. *paratuberculosis* with *Mycobacterium avium* subsp. *avium* Reveals Potential Diagnostic Sequences. *J Clin Microbiol* **40**, 1303–1310 (2002).
135. Rajeev, S., Zhang, Y., Sreevatsan, S., Motiwala, A. S. & Byrum, B. Evaluation of multiple genomic targets for identification and confirmation of *Mycobacterium avium* subsp. *paratuberculosis* isolates using real-time PCR. *Vet Microbiol* **105**, 215–221 (2005).

136. Sibley, J. A., Woodbury, M. R., Appleyard, G. D. & Elkin, B. *Mycobacterium avium* subspecies *paratuberculosis* in Bison (Bison bison) from Northern Canada. *J Wildl Dis* **43**, 775–779 (2007).

137. Crossley, B. M., Zagmutt-Vergara, F. J., Fyock, T. L., Whitlock, R. H. & Gardner, I. A. Fecal shedding of *Mycobacterium avium* subsp. *paratuberculosis* by dairy cows. *Veterinary Microbiology* **107**, 257–263 (2005).

138. Idris, S. M. *et al.* Paratuberculosis: The Hidden Killer of Small Ruminants. *Animals (Basel)* **12**, 12 (2021).

139. Koets, A. P. *et al.* Genetic Variation of Susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* Infection in Dairy Cattle. *Journal of Dairy Science* **83**, 2702–2708 (2000).

140. Macindoe, R. F. H. What can we do about Johne's disease? *Australian Veterinary Journal* **26**, 216–219 (1950).

141. Mizzi, R. *et al.* Comparative Genomics of *Mycobacterium avium* subspecies *paratuberculosis* Sheep Strains. *Frontiers in Veterinary Science* **8**, (2021).

142. Motiwala, A. S. *et al.* Molecular Epidemiology of *Mycobacterium avium* subsp. *paratuberculosis* Isolates Recovered from Wild Animal Species. *J Clin Microbiol* **42**, 1703–1712 (2004).

143. Seaman, J. T. & Thompson, D. R. Johne's disease in sheep. *Australian Veterinary Journal* **61**, 227–229 (1984).

144. van Roermund, H. J. W., Bakker, D., Willemsen, P. T. J. & de Jong, M. C. M. Horizontal transmission of *Mycobacterium avium* subsp. *paratuberculosis* in cattle

in an experimental setting: Calves can transmit the infection to other calves. *Veterinary Microbiology* **122**, 270–279 (2007).

145. Windsor, P. A. Paratuberculosis in sheep and goats. *Vet Microbiol* **181**, 161–169 (2015).
146. Buergelt, C. D. & Ginn, P. E. The histopathologic diagnosis of subclinical Johne's disease in North American bison (*Bison bison*). *Vet Microbiol* **77**, 325–331 (2000).
147. Chiodini, R. J. & Van Kruiningen, H. J. Eastern white-tailed deer as a reservoir of ruminant paratuberculosis. *J Am Vet Med Assoc* **182**, 168–169 (1983).
148. Cook, W. E., Cornish, T. E., Shideler, S., Lasley, B. & Collins, M. T. Radiometric culture of *Mycobacterium avium paratuberculosis* from the feces of tule elk. *J Wildl Dis* **33**, 635–637 (1997).
149. Machackova-Kopecna, M. *et al.* Paratuberculosis and avian tuberculosis infections in one red deer farm studied by *IS900* and *IS901* RFLP analysis. *Vet Microbiol* **105**, 261–268 (2005).
150. Manning, E. J., Kucera, T. E., Gates, N. B., Woods, L. M. & Fallon-McKnight, M. Testing for *Mycobacterium avium* subsp. *paratuberculosis* infection in asymptomatic free-ranging tule elk from an infected herd. *J Wildl Dis* **39**, 323–328 (2003).
151. Beard, P. M. *et al.* Paratuberculosis infection of nonruminant wildlife in Scotland. *J Clin Microbiol* **39**, 1517–1521 (2001).
152. Corn, J. L., Manning, E. J. B., Sreevatsan, S. & Fischer, J. R. Isolation of *Mycobacterium avium* subsp. *paratuberculosis* from free-ranging birds and mammals on livestock premises. *Appl Environ Microbiol* **71**, 6963–6967 (2005).

153. Greig, A. *et al.* Paratuberculosis in wild rabbits (*Oryctolagus cuniculus*). *Vet Rec* **140**, 141–143 (1997).

154. McClure, H. M. *et al. Mycobacterium paratuberculosis* infection in a colony of stumptail macaques (*Macaca arctoides*). *J Infect Dis* **155**, 1011–1019 (1987).

155. Singh, S. V., Singh, A. V., Singh, P. K., Kumar, A. & Singh, B. Molecular identification and characterization of *Mycobacterium avium* subspecies *paratuberculosis* in free living non-human primate (*Rhesus macaques*) from North India. *Comp Immunol Microbiol Infect Dis* **34**, 267–271 (2011).

156. Magombedze, G., Ngonghala, C. N. & Lanzas, C. Evalution of the "Iceberg Phenomenon" in Johne's Disease through Mathematical Modelling. *PLOS ONE* **8**, e76636 (2013).

157. Tiwari, A., VanLeeuwen, J. A., McKenna, S. L. B., Keefe, G. P. & Barkema, H. W. Johne's disease in Canada: Part I: Clinical symptoms, pathophysiology, diagnosis, and prevalence in dairy herds. *Can Vet J* **47**, 874–882 (2006).

158. Whittington, R. & Sergeant, E. Progress towards understanding the spread, detection and control of *Mycobacterium avium* subsp. *paratuberculosis* in animal populations. *Australian Veterinary Journal* **79**, 267–278 (2001).

159. Mortier, R. A., Barkema, H. W., Orsel, K., Wolf, R. & De Buck, J. Shedding patterns of dairy calves experimentally infected with *Mycobacterium avium* subspecies *paratuberculosis*. *Vet Res* **45**, 71 (2014).

160. Waters, W. R. *et al.* Early Induction of Humoral and Cellular Immune Responses during Experimental *Mycobacterium avium* subsp. *paratuberculosis* Infection of Calves. *Infection and Immunity* **71**, 5130–5138 (2003).

161. Eisenberg, S. W. F., Nielen, M. & Koets, A. P. Within-farm transmission of bovine paratuberculosis: recent developments. *Veterinary Quarterly* **32**, 31–35 (2012).

162. Sweeney, R. W. Transmission of Paratuberculosis. *Veterinary Clinics of North America: Food Animal Practice* **12**, 305–312 (1996).

163. Windsor, P. A. & Whittington, R. J. Evidence for age susceptibility of cattle to Johne's disease. *The Veterinary Journal* **184**, 37–44 (2010).

164. Nielsen, S. S., Bjerre, H. & Toft, N. Colostrum and milk as risk factors for infection with *Mycobacterium avium* subspecies *paratuberculosis* in dairy cattle. *J Dairy Sci* **91**, 4610–4615 (2008).

165. Streeter, R. N., Hoffsis, G. F., Bech-Nielsen, S., Shulaw, W. P. & Rings, D. M. Isolation of *Mycobacterium paratuberculosis* from colostrum and milk of subclinically infected cows. *Am J Vet Res* **56**, 1322–1324 (1995).

166. Seitz, S. E. *et al.* Bovine fetal infection with *Mycobacterium paratuberculosis*. *J Am Vet Med Assoc* **194**, 1423–1426 (1989).

167. Sweeney, R. W., Whitlock, R. H. & Rosenberger, A. E. *Mycobacterium paratuberculosis* isolated from fetuses of infected cows not manifesting signs of the disease. *Am J Vet Res* **53**, 477–480 (1992).

168. van Kooten, H. C. J., Mackintosh, C. G. & Koets, A. P. Intra-uterine transmission of paratuberculosis (Johne's disease) in farmed red deer. *N Z Vet J* **54**, 16–20 (2006).

169. Whittington, R. J. & Windsor, P. A. In utero infection of cattle with M*ycobacterium avium* subsp. *paratuberculosis*: a critical review and meta-analysis. *Vet J* **179**, 60–69 (2009).

170. Fecteau, M.-E., Whitlock, R. H., Buergelt, C. D. & Sweeney, R. W. Exposure of young dairy cattle to *Mycobacterium avium* subsp. *paratuberculosis* (MAP) through intensive grazing of contaminated pastures in a herd positive for Johne's disease. *Can Vet J* **51**, 198–200 (2010).

171. Marcé, C., Ezanno, P., Seegers, H., Pfeiffer, D. U. & Fourichon, C. Within-herd contact structure and transmission of *Mycobacterium avium* subspecies *paratuberculosis* in a persistently infected dairy cattle herd. *Preventive Veterinary Medicine* **100**, 116–125 (2011).

172. Whitlock, R. H. & Buergelt, C. Preclinical and Clinical Manifestations of Paratuberculosis (Including Pathology). *Veterinary Clinics of North America: Food Animal Practice* **12**, 345–356 (1996).

173. Hutchinson, L. J. Economic Impact of Paratuberculosis. *Veterinary Clinics of North America: Food Animal Practice* **12**, 373–381 (1996).

174. Bates, A., O'Brien, R., Liggett, S. & Griffin, F. The effect of sub-clinical infection with *Mycobacterium avium* subsp. *paratuberculosis* on milk production in a New Zealand dairy herd. *BMC Veterinary Research* **14**, 93 (2018).

175. Barkema, H. W. *et al.* Knowledge gaps that hamper prevention and control of *Mycobacterium avium* subspecies *paratuberculosis* infection. *Transbound Emerg Dis* **65 Suppl 1**, 125–148 (2018).

176. Ozsvari, L. *et al.* The Impact of Paratuberculosis on Milk Production, Fertility, and Culling in Large Commercial Hungarian Dairy Herds. *Front Vet Sci* **7**, 565324 (2020).

177. Navarro-Gonzalez, N. *et al.* Longitudinal study of *Mycobacterium avium* ssp. *paratuberculosis* fecal shedding patterns and concurrent serological patterns in naturally infected dairy cattle. *Journal of Dairy Science* **102**, 9117–9137 (2019).

178. Fock-Chow-Tho, D., Topp, E., Ibeagha-Awemu, E. A. & Bissonnette, N. Comparison of commercial DNA extraction kits and quantitative PCR systems for better sensitivity in detecting the causative agent of paratuberculosis in dairy cow fecal samples. *J Dairy Sci* **100**, 572–581 (2017).

179. Karuppusamy, S. *et al.* Detection of *Mycobacterium avium* subspecies *paratuberculosis* (MAP) Microorganisms Using Antigenic MAP Cell Envelope Proteins. *Frontiers in Veterinary Science* **8**, (2021).

180. Gilardoni, L. R., Paolicchi, F. A. & Mundo, S. L. Bovine paratuberculosis: a review of the advantages and disadvantages of different diagnostic tests. *Rev Argent Microbiol* **44**, 201–215 (2012).

181. Cheng, Z. *et al.* Characteristics and Epidemiological Investigation of Paratuberculosis in Dairy Cattle in Tai'an, China. *Biomed Res Int* **2020**, 3896754 (2020).

182. Cocito, C. *et al.* Paratuberculosis. *Clinical Microbiology Reviews* **7**, 328–345 (1994).

183. Chaubey, K. K. *et al.* Trends and advances in the diagnosis and control of paratuberculosis in domestic livestock. *Veterinary Quarterly* **36**, 203–227 (2016).

184. Whittington, R. *et al.* Control of paratuberculosis: who, why and how. A review of 48 countries. *BMC Veterinary Research* **15**, 198 (2019).

185. Sweeney, R. w., Collins, M. t., Koets, A. p., McGuirk, S. m. & Roussel, A. j. Paratuberculosis (Johne's Disease) in Cattle and Other Susceptible Species. *Journal of Veterinary Internal Medicine* **26**, 1239–1250 (2012).

186. Arsenault, R. J. *et al.* From mouth to macrophage: mechanisms of innate immune subversion by *Mycobacterium avium* subsp. *paratuberculosis*. *Vet Res* **45**, 54 (2014).

187. Bannantine, J. P. & Bermudez, L. E. No Holes Barred: Invasion of the Intestinal Mucosa by *Mycobacterium avium* subsp. *paratuberculosis*. *Infect Immun* **81**, 3960–3965 (2013).

188. Kobayashi, N., Takahashi, D., Takano, S., Kimura, S. & Hase, K. The Roles of Peyer's Patches and Microfold Cells in the Gut Immune System: Relevance to Autoimmune Diseases. *Frontiers in Immunology* **10**, (2019).

189. Bermudez, L. E., Petrofsky, M., Sommer, S. & Barletta, R. G. Peyer's Patch-Deficient Mice Demonstrate That *Mycobacterium avium* subsp. *paratuberculosis* Translocates across the Mucosal Barrier via both M Cells and Enterocytes but Has Inefficient Dissemination. *Infect Immun* **78**, 3570–3577 (2010).

190. Pott, J. *et al.* Internalization-dependent recognition of *Mycobacterium avium* ssp. *paratuberculosis* by intestinal epithelial cells. *Cellular Microbiology* **11**, 1802–1815 (2009).

191. Kraehenbuhl, J.-P. & Neutra, M. R. Epithelial M Cells: Differentiation and Function. *Annual Review of Cell and Developmental Biology* **16**, 301–332 (2000).

192. Sansonetti, P. J. & Phalipon, A. M cells as ports of entry for enteroinvasive pathogens: Mechanisms of interaction, consequences for the disease process. *Seminars in Immunology* **11**, 193–203 (1999).

193. Kuo, C.-J., Bell, H., Hsieh, C.-L., Ptak, C. P. & Chang, Y.-F. Novel Mycobacteria Antigen 85 Complex Binding Motif on Fibronectin. *J Biol Chem* **287**, 1892–1902 (2012).

194. Secott, T. E., Lin, T. L. & Wu, C. C. Fibronectin Attachment Protein Homologue Mediates Fibronectin Binding by *Mycobacterium avium* subsp. *paratuberculosis*. *Infect Immun* **69**, 2075–2082 (2001).

195. Sigurðardóttir, Ó. G., Valheim, M. & Press, C. M. Establishment of *Mycobacterium avium* subsp. *paratuberculosis* infection in the intestine of ruminants. *Advanced Drug Delivery Reviews* **56**, 819–834 (2004).

196. Khare, S. *et al.* Early Phase Morphological Lesions and Transcriptional Responses of Bovine Ileum Infected with *Mycobacterium avium* subsp. *paratuberculosis*. *Vet Pathol* **46**, 717–728 (2009).

197. Byun, E.-H. *et al. Mycobacterium paratuberculosis* CobT Activates Dendritic Cells via Engagement of Toll-like Receptor 4 Resulting in Th1 Cell Expansion. *J Biol Chem* **287**, 38609–38624 (2012).

198. Lee, J. S. *et al. Mycobacterium avium* subsp. *paratuberculosis* Fibronectin Attachment Protein Activates Dendritic Cells and Induces a Th1 Polarization. *Infection and Immunity* **77**, 2979–2988 (2009).

199. Lei, L. & Hostetter, J. M. Limited phenotypic and functional maturation of bovine monocyte-derived dendritic cells following *Mycobacterium avium* subspecies *paratuberculosis* infection in vitro. *Veterinary Immunology and Immunopathology* **120**, 177–186 (2007).

200. Momotani, E., Whipple, D. L., Thiermann, A. B. & Cheville, N. F. Role of M Cells and Macrophages in the Entrance of *Mycobacterium paratuberculosis* into Domes of Ileal Peyer's Patches in Calves. *Vet Pathol* **25**, 131–137 (1988).

201. Guirado, E., Schlesinger, L. S. & Kaplan, G. Macrophages in Tuberculosis: Friend or Foe. *Semin Immunopathol* **35**, 563–583 (2013).

202. de Chastellier, C. The many niches and strategies used by pathogenic mycobacteria for survival within host macrophages. *Immunobiology* **214**, 526–542 (2009).

203. Hmama, Z., Peña-Díaz, S., Joseph, S. & Av-Gay, Y. Immunoevasion and immunosuppression of the macrophage by *Mycobacterium tuberculosis*. *Immunological Reviews* **264**, 220–232 (2015).

204. Koul, A., Herget, T., Klebl, B. & Ullrich, A. Interplay between mycobacteria and host signalling pathways. *Nat Rev Microbiol* **2**, 189–202 (2004).

205. Rumsey, J., Valentine, J. & Naser, S. Inhibition of phagosome maturation and survival of *Mycobacterium avium* subspecies *paratuberculosis* in polymorphonuclear leukocytes from Crohn's disease patients. *Medical science monitor: international medical journal of experimental and clinical research* **12**, BR130-9 (2006).

206. Garvey, M. *Mycobacterium avium paratuberculosis*: A Disease Burden on the Dairy Industry. *Animals (Basel)* **10**, 1773 (2020).

207. Fernández, M. *et al.* Macrophage Subsets Within Granulomatous Intestinal Lesions in Bovine Paratuberculosis. *Vet Pathol* **54**, 82–93 (2017).

208. González, J. *et al.* Histopathological Classification of Lesions associated with Natural Paratuberculosis Infection in Cattle. *Journal of Comparative Pathology* **133**, 184–196 (2005).

209. Koets, A. P., Eda, S. & Sreevatsan, S. The within host dynamics of *Mycobacterium avium* ssp. *paratuberculosis* infection in cattle: where time and place matter. *Veterinary Research* **46**, 61 (2015).

210. Catton, B. A. Paucibacillary paratuberculosis in a goat. *Can Vet J* **43**, 787–788 (2002).

211. Clark, R., Griffin, J. & Mackintosh, C. Johne's disease caused by *Mycobacterium avium* subsp. *paratuberculosis* infection in red deer (*Cervus elaphus*): An histopathological grading system, and comparison of paucibacillary and multibacillary disease. *New Zealand Veterinary Journal* **58**, 90–97 (2010).

212. Smeed, J. A., Watkins, C. A., Rhind, S. M. & Hopkins, J. Differential cytokine gene expression profiles in the three pathological forms of sheep paratuberculosis. *BMC Vet Res* **3**, 18 (2007).

213. Chaitanya, R. K., Lakshmi Kavitha, K. & Sreedevi, B. Current diagnostic approaches for Paratuberculosis- A Review. *JLivestSci* **13**, 201–207 (2022).

214. Gardner, I. A. *et al.* Consensus-based reporting standards for diagnostic test accuracy studies for paratuberculosis in ruminants. *Prev Vet Med* **101**, 18–34 (2011).

215. Alonso, B. Chapter 3.1.15. Paratuberculosis (Johne's Disease). in *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals* (World Organization for Animal Health (OIE), 2023).

216. Coelho, A. C. *et al.* Comparative evaluation of PCR in Ziehl-Neelsen stained smears and PCR in tissues for diagnosis of *Mycobacterium avium* subsp. *paratuberculosis*. *Indian J Exp Biol* **48**, 948–950 (2010).

217. Weber, M. F., Verhoeff, J., van Schaik, G. & van Maanen, C. Evaluation of Ziehl-Neelsen stained faecal smear and ELISA as tools for surveillance of clinical paratuberculosis in cattle in the Netherlands. *Prev Vet Med* **92**, 256–266 (2009).

218. Nielsen, S. S. & Toft, N. Ante mortem diagnosis of paratuberculosis: A review of accuracies of ELISA, interferon-γ assay and faecal culture techniques. *Veterinary Microbiology* **129**, 217–235 (2008).

219. Whitlock, R. H., Wells, S. J., Sweeney, R. W. & Van Tiem, J. ELISA and fecal culture for paratuberculosis (Johne's disease): sensitivity and specificity of each method. *Veterinary Microbiology* **77**, 387–398 (2000).

220. Luttikholt, S. *et al. Mycobacterium avium* subsp. *paratuberculosis* ELISA Responses in Milk Samples from Vaccinated and Nonvaccinated Dairy Goat Herds in The Netherlands. *Vet Sci* **6**, 58 (2019).

221. Lavers, C. J., Barkema, H. W., Dohoo, I. R., McKenna, S. L. B. & Keefe, G. P. Evaluation of milk ELISA for detection of *Mycobacterium avium* subspecies *paratuberculosis* in dairy herds and association with within-herd prevalence. *J Dairy Sci* **97**, 299–309 (2014).

222. Gumber, S., Eamens, G. & Whittington, R. J. Evaluation of a Pourquier ELISA kit in relation to agar gel immunodiffusion (AGID) test for assessment of the humoral immune response in sheep and goats with and without *Mycobacterium paratuberculosis* infection. *Vet Microbiol* **115**, 91–101 (2006).

223. Kalis, C. H. J., Barkema, H. W., Hesselink, J. W., van Maanen, C. & Collins, M. T. Evaluation of two absorbed enzyme-linked immunosorbent assays and a complement fixation test as replacements for fecal culture in the detection of cows shedding *Mycobacterium avium* subspecies *paratuberculosis*. *J Vet Diagn Invest* **14**, 219–224 (2002).

224. Whittington, R. J., Marsh, I. B. & Reddacliff, L. A. Survival of *Mycobacterium avium* subsp. *paratuberculosis* in Dam Water and Sediment. *Appl Environ Microbiol* **71**, 5304–5308 (2005).

225. Whittington, R. J., Marshall, D. J., Nicholls, P. J., Marsh, I. B. & Reddacliff, L. A. Survival and Dormancy of *Mycobacterium avium* subsp. *paratuberculosis* in the Environment. *Appl Environ Microbiol* **70**, 2989–3004 (2004).

226. Raizman, E. A. *et al.* Leaching of *Mycobacterium avium* subsp. *paratuberculosis* in Soil under *In Vitro* Conditions. *Veterinary Medicine International* **2011**, e506239 (2011).

227. Salgado, M. *et al.* Fate of *Mycobacterium avium* subsp. *paratuberculosis* after Application of Contaminated Dairy Cattle Manure to Agricultural Soils. *Appl Environ Microbiol* **77**, 2122–2129 (2011).

228. Samba-Louaka, A. *et al.* Environmental *Mycobacterium avium* subsp. *paratuberculosis* Hosted by Free-Living Amoebae. *Front Cell Infect Microbiol* **8**, 28 (2018).

229. Whan, L., Grant, I. R. & Rowe, M. T. Interaction between *Mycobacterium avium* subsp. *paratuberculosis* and environmental protozoa. *BMC Microbiol* **6**, 63 (2006).

230. Cirillo, J. D., Falkow, S., Tompkins, L. S. & Bermudez, L. E. Interaction of *Mycobacterium avium* with environmental amoebae enhances virulence. *Infect Immun* **65**, 3759–3767 (1997).

231. Miltner, E. C. & Bermudez, L. E. *Mycobacterium avium* Grown in *Acanthamoeba castellanii* Is Protected from the Effects of Antimicrobials. *Antimicrob Agents Chemother* **44**, 1990–1994 (2000).

232. Salgado, M. *et al.* Application of cattle slurry containing *Mycobacterium avium* subsp. *paratuberculosis* (MAP) to grassland soil and its effect on the relationship between MAP and free-living amoeba. *Veterinary Microbiology* **175**, 26–34 (2015).

233. Mura, M. *et al.* Replication and Long-Term Persistence of Bovine and Human Strains of *Mycobacterium avium* subsp. *paratuberculosis* within *Acanthamoeba polyphaga*. *Applied and Environmental Microbiology* **72**, 854–859 (2006).

234. White, C. I., Birtles, R. J., Wigley, P. & Jones, P. H. *Mycobacterium avium* subspecies *paratuberculosis* in free-living amoebae isolated from fields not used for grazing. *Veterinary Record* **166**, 401–402 (2010).

235. Flemming, H.-C. *et al.* Biofilms: an emergent form of bacterial life. *Nat Rev Microbiol* **14**, 563–575 (2016).

236. Ojha, A. K. *et al.* Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. *Mol Microbiol* **69**, 164–174 (2008).

237. Recht, J., Martínez, A., Torello, S. & Kolter, R. Genetic Analysis of Sliding Motility in *Mycobacterium smegmatis*. *J Bacteriol* **182**, 4348–4351 (2000).

238. Recht, J. & Kolter, R. Glycopeptidolipid Acetylation Affects Sliding Motility and Biofilm Formation in *Mycobacterium smegmatis*. *J Bacteriol* **183**, 5718–5724 (2001).

239. Yamazaki, Y., Danelishvili, L., Wu, M., MacNab, M. & Bermudez, L. E. *Mycobacterium avium* Genes Associated with the Ability To Form a Biofilm. *Appl Environ Microbiol* **72**, 819–825 (2006).

240. Pang, L., Tian, X., Pan, W. & Xie, J. Structure and function of mycobacterium glycopeptidolipids from comparative genomics perspective. *J Cell Biochem* **114**, 1705–1713 (2013).

241. Schorey, J. S. & Sweet, L. The mycobacterial glycopeptidolipids: structure, function, and their role in pathogenesis. *Glycobiology* **18**, 832–841 (2008).

242. Ssekitoleko, J. *et al. Mycobacterium avium* subsp. *paratuberculosis* Virulence: A Review. *Microorganisms* **9**, 2623 (2021).

243. Wu, C. *et al.* A Novel Cell Wall Lipopeptide Is Important for Biofilm Formation and Pathogenicity of *Mycobacterium avium* subspecies *paratuberculosis*. *Microb Pathog* **46**, 222–230 (2009).

244. Beumer, A. *et al.* Detection of *Mycobacterium avium* subsp. *paratuberculosis* in Drinking Water and Biofilms by Quantitative PCR. *Applied and Environmental Microbiology* **76**, 7367–7370 (2010).

245. Cook, K. L., Britt, J. S. & Bolster, C. H. Survival of *Mycobacterium avium* subsp. *paratuberculosis* in biofilms on livestock watering trough materials. *Veterinary Microbiology* **141**, 103–109 (2010).

246. Lehtola, M. J., Torvinen, E., Miettinen, I. T. & Keevil, C. W. Fluorescence In Situ Hybridization Using Peptide Nucleic Acid Probes for Rapid Detection of *Mycobacterium avium* subsp. *avium* and *Mycobacterium avium* subsp. *paratuberculosis* in Potable-Water Biofilms. *Applied and Environmental Microbiology* **72**, 848–853 (2006).

247. Lamont, E. A., Bannantine, J. P., Armién, A., Ariyakumar, D. S. & Sreevatsan, S. Identification and Characterization of a Spore-Like Morphotype in Chronically Starved *Mycobacterium avium* subsp. *paratuberculosis* Cultures. *PLOS ONE* **7**, e30648 (2012).

248. Ghosh, J. *et al.* Sporulation in mycobacteria. *Proceedings of the National Academy of Sciences* **106**, 10781–10786 (2009).

249. Santema, W., Rutten, V. & Koets, A. Bovine paratuberculosis: recent advances in vaccine development. *Veterinary Quarterly* **31**, 183–191 (2011).

250. Coad, M., Clifford, D. J., Vordermeier, H. M. & Whelan, A. O. The consequences of vaccination with the Johne's disease vaccine, Gudair, on diagnosis of bovine tuberculosis. *Vet Rec* **172**, 266 (2013).

251. Muskens, J., van Zijderveld, F., Eger, A. & Bakker, D. Evaluation of the long-term immune response in cattle after vaccination against paratuberculosis in two Dutch dairy herds. *Vet Microbiol* **86**, 269–278 (2002).

252. Serrano, M. *et al.* Tuberculosis Detection in Paratuberculosis Vaccinated Calves: New Alternatives against Interference. *PLoS One* **12**, e0169735 (2017).

253. Ghosh, P., Shippy, D. C. & Talaat, A. M. Superior protection elicited by live-attenuated vaccines in the murine model of paratuberculosis. *Vaccine* **33**, 7262–7270 (2015).

254. Park, K. T. *et al.* Evaluation of two mutants of *Mycobacterium avium* subsp. *paratuberculosis* as candidates for a live attenuated vaccine for Johne's disease. *Vaccine* **29**, 4709–4719 (2011).

255. Shippy, D. C. *et al.* Superior Protection from Live-Attenuated Vaccines Directed against Johne's Disease. *Clin Vaccine Immunol* **24**, e00478-16 (2017).

256. Koets, A. *et al.* Mycobacterial 70 kD heat-shock protein is an effective subunit vaccine against bovine paratuberculosis. *Vaccine* **24**, 2550–2559 (2006).

257. Rosseels, V. & Huygen, K. Vaccination against paratuberculosis. *Expert Review of Vaccines* **7**, 817–832 (2008).

258. Thakur, A., Aagaard, C., Stockmarr, A., Andersen, P. & Jungersen, G. Cell-Mediated and Humoral Immune Responses after Immunization of Calves with a Recombinant Multiantigenic *Mycobacterium avium* subsp. *paratuberculosis* Subunit Vaccine at Different Ages. *Clinical and Vaccine Immunology* **20**, 551–558 (2013).

259. Rasmussen, P., Barkema, H. W., Beaulieu, E., Mason, S. & Hall, D. C. Estimation of the value of Johne's disease (paratuberculosis) control to Canadian dairy producers. *Preventive Veterinary Medicine* **189**, 105297 (2021).

260. Lu, Z. *et al.* The importance of culling in Johne's disease control. *Journal of Theoretical Biology* **254**, 135–146 (2008).

261. McKenna, S. L. B. *et al.* Proposed Canadian Voluntary National Johne's Disease Prevention and Control Program. *Can Vet J* **47**, 539–541 (2006).

262. Elliott, G. N., Hough, R. L., Avery, L. M., Maltin, C. A. & Campbell, C. D. Environmental risk factors in the incidence of Johne's disease. *Critical Reviews in Microbiology* **41**, 488–507 (2015).

263. Benedictus, A. *et al.* Transmission parameters of *Mycobacterium avium* subspecies *paratuberculosis* infections in a dairy herd going through a control program. *Prev Vet Med* **83**, 215–227 (2008).

264. Collins, M. T., Eggleston, V. & Manning, E. J. B. Successful control of Johne's disease in nine dairy herds: Results of a six-year field trial. *Journal of Dairy Science* **93**, 1638–1643 (2010).

265. Ferrouillet, C., Wells, S. J., Hartmann, W. L., Godden, S. M. & Carrier, J. Decrease of Johne's disease prevalence and incidence in six Minnesota, USA, dairy cattle herds on a long-term management program. *Preventive Veterinary Medicine* **88**, 128–137 (2009).

266. Donat, K. The Thuringian bovine paratuberculosis control programme – results and experiences. *Berl Münch Tierärztl Wochensch* .-.. (2016) doi:10.2376/0005-9366-129-15129.

267. Kalis, C. H. J., Collins, M. T., Barkema, H. W. & Hesselink, J. W. Certification of herds as free of *Mycobacterium paratuberculosis* infection: actual pooled faecal results versus certification model predictions. *Preventive Veterinary Medicine* **65**, 189–204 (2004).

268. McKenna, S. L. B., Keefe, G. P., Tiwari, A., VanLeeuwen, J. & Barkema, H. W. Johne's disease in Canada Part II: Disease impacts, risk factors, and control programs for dairy producers. *Can Vet J* **47**, 1089–1099 (2006).

269. Hodgeman, R. *et al.* Molecular characterisation of *Mycobacterium avium* subsp. *paratuberculosis* in Australia. *BMC Microbiology* **21**, 101 (2021).

270. Sohal, J. S. *et al.* Molecular characterization of *Mycobacterium avium* subspecies *paratuberculosis* C-type and S-type isolated from sheep and goats by using a combination of MIRU-VNTR loci. *Can J Vet Res* **83**, 160–167 (2019).

271. Szteyn, J., Liedtke, K., Wiszniewska-Łaszczych, A., Wysok, B. & Wojtacka, J. Isolation and molecular typing of *Mycobacterium avium* subsp. *paratuberculosis* from faeces of dairy cows. *Pol J Vet Sci* **23**, 415–422 (2020).

272. Biet, F. *et al.* Inter- and Intra-subtype genotypic differences that differentiate *Mycobacterium avium* subspecies *paratuberculosis* strains. *BMC Microbiology* **12**, 264 (2012).

273. Semret, M., Turenne, C. Y., de Haas, P., Collins, D. M. & Behr, M. A. Differentiating Host-Associated Variants of *Mycobacterium avium* by PCR for Detection of Large Sequence Polymorphisms. *Journal of Clinical Microbiology* **44**, 881–887 (2006).

274. Collins, D. M., Gabric, D. M. & de Lisle, G. W. Identification of two groups of *Mycobacterium paratuberculosis* strains by restriction endonuclease analysis and DNA hybridization. *J Clin Microbiol* **28**, 1591–1596 (1990).

275. de Juan, L., Mateos, A., Domínguez, L., Sharp, J. M. & Stevenson, K. Genetic diversity of *Mycobacterium avium* subspecies *paratuberculosis* isolates from goats detected by pulsed-field gel electrophoresis. *Vet Microbiol* **106**, 249–257 (2005).

276. Whittington, R. J., Marsh, I. B. & Whitlock, R. H. Typing of *IS 1311* polymorphisms confirms that bison (*Bison bison*) with paratuberculosis in Montana are infected with

a strain of *Mycobacterium avium* subsp. *paratuberculosis* distinct from that occurring in cattle and other domesticated livestock. *Molecular and cellular probes* **15**, (2001).

277. Marsh, I., Whittington, R. & Cousins, D. PCR-restriction endonuclease analysis for identification and strain typing of *Mycobacterium avium* subsp. *paratuberculosis* and *Mycobacterium avium* subsp. *avium* based on polymorphisms in *IS1311*. *Molecular and Cellular Probes* **13**, 115–126 (1999).
278. Pavlik, I. *et al.* Standardisation of restriction fragment length polymorphism analysis for *Mycobacterium avium* subspecies *paratuberculosis*. *Journal of Microbiological Methods* **38**, 155–167 (1999).
279. Stevenson, K. *et al.* Molecular Characterization of Pigmented and Nonpigmented Isolates of *Mycobacterium avium* subsp. *paratuberculosis*. *Journal of Clinical Microbiology* **40**, 1798–1804 (2002).
280. Yadav, D. *et al.* Pathogenic 'Bison-type' *Mycobacterium avium* subspecies *paratuberculosis* genotype characterized from riverine buffalo (*Bubalus bubalis*) in North India. *Comparative Immunology, Microbiology and Infectious Diseases* **31**, 373–387 (2008).
281. Sohal, J. S., Singh, S. V., Singh, P. K., Singh, A. V. & Kumar, N. A new marker *IS1311* L2 PCR-REA for identification of 'Indian Bison' type *Mycobacterium avium* subspecies *paratuberculosis*. *Indian Journal of Biotechnology* **12**, 204–207 (2013).
282. Sevilla, I. x *et al.* Molecular typing of *Mycobacterium avium* subspecies *paratuberculosis* strains from different hosts and regions. *Rev Sci Tech* **24**, 1061–1066 (2005).

283. Fernández, M. *et al.* Experimental infection of lambs with C and S-type strains of *Mycobacterium avium* subspecies *paratuberculosis*: immunological and pathological findings. *Veterinary Research* **45**, 5 (2014).

284. Stevenson, K. *et al.* Occurrence of *Mycobacterium avium* subspecies *paratuberculosis* across host species and European countries with evidence for transmission between wildlife and domestic ruminants. *BMC Microbiol* **9**, 212 (2009).

285. Verdugo, C. *et al.* Molecular epidemiology of *Mycobacterium avium* subsp. *paratuberculosis* isolated from sheep, cattle and deer on New Zealand pastoral farms. *Prev Vet Med* **117**, 436–446 (2014).

286. Amonsin, A. *et al.* Multilocus Short Sequence Repeat Sequencing Approach for Differentiating among *Mycobacterium avium* subsp. *paratuberculosis* Strains. *J Clin Microbiol* **42**, 1694–1702 (2004).

287. Thibault, V. C. *et al.* New Variable-Number Tandem-Repeat Markers for Typing *Mycobacterium avium* subsp. *paratuberculosis* and *M. avium* Strains: Comparison with *IS900* and *IS1245* Restriction Fragment Length Polymorphism Typing. *Journal of Clinical Microbiology* **45**, 2404–2410 (2007).

288. Conde, C. *et al.* Genetic Features of *Mycobacterium avium* subsp. *paratuberculosis* Strains Circulating in the West of France Deciphered by Whole-Genome Sequencing. *Microbiol Spectr* **10**, e0339222 (2022).

289. Davidson, F. W., Ahlstrom, C., De Buck, J., Whitney, H. G. & Tahlan, K. Examination of *Mycobacterium avium* subspecies *paratuberculosis* mixed genotype

infections in dairy animals using a whole genome sequencing approach. *PeerJ* **4**, e2793 (2016).

290. Perets, V. *et al.* Evidence for local and international spread of *Mycobacterium avium* subspecies *paratuberculosis* through whole genome sequencing of isolates from the island of Ireland. *Vet Microbiol* **268**, 109416 (2022).
291. Adami, A. J. & Cervantes, J. L. The Microbiome at the Pulmonary Alveolar Niche: How It Affects the Human Innate Response against *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* **95**, 651–658 (2015).
292. Namasivayam, S. *et al.* Correlation between Disease Severity and the Intestinal Microbiome in *Mycobacterium tuberculosis*-Infected Rhesus Macaques. *mBio* **10**, e01018-19 (2019).
293. Namasivayam, S. *et al.* Patients infected with *Mycobacterium africanum* versus *Mycobacterium tuberculosis* possess distinct intestinal microbiota. *PLoS Negl Trop Dis* **14**, e0008230 (2020).
294. Cohen, T. *et al.* Mixed-strain *Mycobacterium tuberculosis* infections and the implications for tuberculosis treatment and control. *Clin Microbiol Rev* **25**, 708–719 (2012).
295. McNaughton, A. L. *et al.* Prevalence of mixed genotype hepatitis C virus infections in the UK as determined by genotype-specific PCR and deep sequencing. *J Viral Hepat* **25**, 524–534 (2018).
296. Taylor, L. H., Walliker, D. & Read, A. F. Mixed-genotype infections of malaria parasites: within-host dynamics and transmission success of competing clones. *Proc Biol Sci* **264**, 927–935 (1997).

297. Feil, E. J. Small change: keeping pace with microevolution. *Nat Rev Microbiol* **2**, 483–495 (2004).

298. Jordan, I. K., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Microevolutionary Genomics of Bacteria. *Theoretical Population Biology* **61**, 435–447 (2002).

299. Ley, S. D., de Vos, M., Van Rie, A. & Warren, R. M. Deciphering Within-Host Microevolution of *Mycobacterium tuberculosis* through Whole-Genome Sequencing: the Phenotypic Impact and Way Forward. *Microbiology and Molecular Biology Reviews* **83**, e00062-18 (2019).

300. Shin, S. S. *et al.* Mixed *Mycobacterium tuberculosis*–Strain Infections Are Associated with Poor Treatment Outcomes Among Patients With Newly Diagnosed Tuberculosis, Independent of Pretreatment Heteroresistance. *The Journal of Infectious Diseases* **218**, 1974 (2018).

301. van Rie, A. *et al.* Reinfection and Mixed Infection Cause Changing *Mycobacterium tuberculosis* Drug-Resistance Patterns. *Am J Respir Crit Care Med* **172**, 636–642 (2005).

302. Gengenbacher, M. & Kaufmann, S. H. E. *Mycobacterium tuberculosis*: success through dormancy. *FEMS Microbiol Rev* **36**, 514–532 (2012).

303. Baldeviano-Vidalón, G. C. *et al.* Multiple infection with resistant and sensitive *M. tuberculosis* strains during treatment of pulmonary tuberculosis patients. *Int J Tuberc Lung Dis* **9**, 1155–1160 (2005).

304. Huang, H.-Y. *et al.* Mixed Infection with Beijing and Non-Beijing Strains and Drug Resistance Pattern of *Mycobacterium tuberculosis*. *J Clin Microbiol* **48**, 4474–4480 (2010).

305. Yoshida, S. *et al.* Association between sequevar and antibiotic treatment outcome in patients with *Mycobacterium abscessus* complex infections in Japan. *J Med Microbiol* **67**, 74–82 (2018).

306. Asare-Baah, M., Séraphin, M. N., Salmon, L. A. T. & Lauzardo, M. Effect of mixed strain infections on clinical and epidemiological features of tuberculosis in Florida. *Infect Genet Evol* **87**, 104659 (2021).

307. Baik, Y. *et al.* Possible Transmission Mechanisms of Mixed *Mycobacterium tuberculosis* Infection in High HIV Prevalence Country, Botswana. *Emerg Infect Dis* **26**, 953–960 (2020).

308. Cohen, T. *et al.* Within-Host Heterogeneity of *Mycobacterium tuberculosis* Infection Is Associated with Poor Early Treatment Response: A Prospective Cohort Study. *The Journal of Infectious Diseases* **213**, 1796–1799 (2016).

309. Fang, R. *et al.* Mixed infections of *Mycobacterium tuberculosis* in tuberculosis patients in Shanghai, China. *Tuberculosis (Edinb)* **88**, 469–473 (2008).

310. Streit, E., Millet, J. & Rastogi, N. *Mycobacterium tuberculosis* polyclonal infections and microevolution identified by MIRU-VNTRs in an epidemiological study. *Int J Mycobacteriol* **4**, 222–227 (2015).

311. Fecteau, M.-E. & Whitlock, R. H. Treatment and chemoprophylaxis for paratuberculosis. *Vet Clin North Am Food Anim Pract* **27**, 547–557, v (2011).

312. Krishnan, M. Y., Manning, E. J. B. & Collins, M. T. Comparison of three methods for susceptibility testing of *Mycobacterium avium* subsp. *paratuberculosis* to 11 antimicrobial drugs. *Journal of Antimicrobial Chemotherapy* **64**, 310–316 (2009).

313. Bryant, J. M. *et al.* Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infectious Diseases* **13**, 110 (2013).

314. Magombedze, G., Shiri, T., Eda, S. & Stabel, J. R. Inferring biomarkers for *Mycobacterium avium* subsp. *paratuberculosis* infection and disease progression in cattle using experimental data. *Sci Rep* **7**, 44765 (2017).

315. Podder, M. P., Banfield, S. E., Keefe, G. P., Whitney, H. G. & Tahlan, K. Typing of *Mycobacterium avium* subspecies *paratuberculosis* Isolates from Newfoundland Using Fragment Analysis. *PLOS ONE* **10**, e0126071 (2015).

316. Adams, L. V. *et al.* Molecular Epidemiology of HIV-Associated Tuberculosis in Dar es Salaam, Tanzania: Strain Predominance, Clustering, and Polyclonal Disease. *J Clin Microbiol* **50**, 2645–2650 (2012).

317. Farmanfarmaei, G. *et al.* Bias in detection of *Mycobacterium tuberculosis* polyclonal infection: Use clinical samples or cultures? *Mol Cell Probes* **33**, 1–3 (2017).

318. Kargarpour Kamakoli, M. *et al.* Evaluation of the impact of polyclonal infection and heteroresistance on treatment of tuberculosis patients. *Sci Rep* **7**, 41410 (2017).

319. Kargarpour Kamakoli, M. *et al.* Application of MIRU–VNTR on smear slides: a shortcut for detection of polyclonal infections in tuberculosis patients. *Mol Biol Rep* **47**, 1681–1689 (2020).

320. Nathavitharana, R. R. *et al.* Polyclonal Pulmonary Tuberculosis Infections and Risk for Multidrug Resistance, Lima, Peru. *Emerg Infect Dis* **23**, 1887–1890 (2017).

321. Bang, D., Rasmussen, E. M. & Andersen, A. B. *Mycobacterium arosiense*, an unexpected cause of osteomyelitis in a patient with sarcoidosis: a case report. *BMC Infectious Diseases* **19**, 994 (2019).

322. Tortoli, E. *et al.* Pulmonary Disease Due to *Mycobacterium arosiense*, an Easily Misidentified Pathogenic Novel *Mycobacterium*. *J Clin Microbiol* **47**, 1947–1949 (2009).

323. Mijs, W. *et al.* Molecular evidence to support a proposal to reserve the designation *Mycobacterium avium* subsp. *avium* for bird-type isolates and '*M. avium* subsp. *hominissuis*' for the human/porcine type of *M. avium*. *International Journal of Systematic and Evolutionary Microbiology* **52**, 1505–1518 (2002).

324. Frothingham, R. & Wilson, K. H. Sequence-based differentiation of strains in the *Mycobacterium avium* complex. *J Bacteriol* **175**, 2818–2825 (1993).

325. Barretto, A. R. *et al.* A fatal case of pulmonary infection by *Mycobacterium colombiense* in Para State, Amazon Region, Brazil. *Diagn Microbiol Infect Dis* **85**, 344–346 (2016).

326. Esparcia, Ó., Navarro, F., Quer, M. & Coll, P. Lymphadenopathy Caused by *Mycobacterium colombiense*. *J Clin Microbiol* **46**, 1885–1887 (2008).

327. Yu, X. & Jiang, W. *Mycobacterium colombiense* and *Mycobacterium avium* Complex Causing Severe Pneumonia in a Patient with HIV Identified by a Novel Molecular-Based Method. *Infect Drug Resist* **14**, 11–16 (2021).

328. Jaworski, R. *et al. Mycobacterium chimaera* – a new threat for cardiac surgical patients? *Kardiochir Torakochirurgia Pol* **14**, 22–26 (2017).

329. Lecorche, E. *et al.* Disseminated *Mycobacterium chimaera* Following Open-Heart Surgery, the Heater–Cooler Unit Worldwide Outbreak: Case Report and Minireview. *Frontiers in Medicine* **7**, (2020).

330. Lu, D. *et al.* Case report of the first cured patient with *Mycobacterium chimaera* infection following cardiac valve replacement in the mainland of China. *Antimicrobial Resistance & Infection Control* **10**, 141 (2021).

331. Miskoff, J. A. & Chaudhri, M. *Mycobacterium chimaera*: A Rare Presentation. *Cureus* **10**, (2018).

332. Rubinstein, M. *et al. Mycobacterium intracellulare* subsp. *chimaera* from Cardio Surgery Heating-Cooling Units and from Clinical Samples in Israel Are Genetically Unrelated. *Pathogens* **10**, 1392 (2021).

333. Castejon, M., Menéndez, M. C., Comas, I., Vicente, A. & Garcia, M. J. Whole-genome sequence analysis of the *Mycobacterium avium* complex and proposal of the transfer of *Mycobacterium yongonense* to *Mycobacterium intracellulare* subsp. *yongonense* subsp. nov. *International Journal of Systematic and Evolutionary Microbiology* **68**, 1998–2005 (2018).

334. Kim, W.-Y. *et al.* Disseminated *Mycobacterium intracellulare* Infection in an Immunocompetent Host. *Tuberc Respir Dis (Seoul)* **72**, 452–456 (2012).

335. Lari, N. & Rindi, L. High genetic heterogeneity of *Mycobacterium intracellulare* isolated from respiratory specimens. *BMC Microbiology* **22**, 5 (2022).

336. Mnyambwa, N. P. *et al.* Genome sequence of *Mycobacterium yongonense* RT 955-2015 isolate from a patient misdiagnosed with multidrug-resistant tuberculosis: First

clinical detection in Tanzania. *International Journal of Infectious Diseases* **71**, 82–88 (2018).

337. Tortoli, E., Mariottini, A., Pierotti, P., Simonetti, T. M. & Rossolini, G. M. *Mycobacterium yongonense* in Pulmonary Disease, Italy. *Emerg Infect Dis* **19**, 1902–1904 (2013).

338. Hase, I. *et al.* Disseminated *Mycobacterium gordonae* and *Mycobacterium mantenii* infection with elevated anti-IFN-γ neutralizing autoantibodies. *Journal of Infection and Chemotherapy* **21**, 468–472 (2015).

339. Mahon, C. & Dalton, S. C. *Mycobacterium mantenii*: a rare cause of chronic ulceration in an immunocompetent female in the southern hemisphere. **134**, (2021).

340. Nebreda Mayoral, T., Andrés Andrés, A. G., Fuentes Carretero, S., Calleja Fernández, R. & Jiménez Pajares, M. S. Cervicofacial lymphadenitis due to *Mycobacterium mantenii*: rapid and reliable identification by MALDI-TOF MS. *New Microbes New Infect* **22**, 1–3 (2017).

341. Azzali, A. *et al.* First case of *Mycobacterium marseillense* lymphadenitis in a child. *Ital J Pediatr* **43**, 92 (2017).

342. Grottola, A. *et al.* Pulmonary Disease Caused by *Mycobacterium marseillense*, Italy. *Emerg Infect Dis* **20**, 1769–1770 (2014).

343. Kim, S.-Y. *et al.* First case of nontuberculous mycobacterial lung disease caused by *Mycobacterium marseillense* in a patient with systemic lupus erythematosus. *Diagnostic Microbiology and Infectious Disease* **79**, 355–357 (2014).

344. Nomura, Y. *et al.* Tenosynovitis caused by *Mycobacterium marseillense*, initially identified as *Mycobacterium avium* complex using AccuProbe and COBAS TaqMan. *BMC Infectious Diseases* **21**, 1092 (2021).

345. Xie, B. *et al. Mycobacterium marseillense* Infection in Human Skin, China, 2018. *Emerging Infectious Diseases* **25**, (2019).

346. Azar, M., Zimbric, M., Shedden, K. & Caverly, L. J. Distribution and outcomes of infection of *Mycobacterium avium* complex species in cystic fibrosis. *J Cyst Fibros* **19**, 232–235 (2020).

347. Zurita, J. *et al.* Characterization of the first report of *Mycobacterium timonense* infecting an HIV patient in an Ecuadorian hospital. *Clinical Microbiology and Infection* **20**, O1113–O1116 (2014).

348. McGowan, J. *et al.* PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology* **75**, 40–46 (2016).

349. Whitlock, R. H. & Rosenberger, A. E. Fecal culture protocol for *Mycobacterium paratuberculosis*. A recommended procedure. *Proc. Annual Meeting US Animal Health Association* **94**, (1990).

350. Whittington, R. J. *et al.* Evaluation of Modified BACTEC 12B Radiometric Medium and Solid Media for Culture of *Mycobacterium avium* subsp. *paratuberculosis* from Sheep. *J Clin Microbiol* **37**, 1077–1083 (1999).

351. Whittington, R. J. *et al.* Development and validation of a liquid medium (M7H9C) for routine culture of *Mycobacterium avium* subsp. *paratuberculosis* to replace modified Bactec 12B medium. *J Clin Microbiol* **51**, 3993–4000 (2013).

352. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

353. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. (2010).

354. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).

355. Conde, C. *et al.* Draft Genome Sequences of 142 *Mycobacterium avium* subsp. *paratuberculosis* Strains Isolated from Naturally Infected Dairy Cattle. *Microbiol Resour Announc* **10**, e0069721 (2021).

356. Lu, J. KrakenTools. (2020).

357. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455–477 (2012).

358. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* **70**, e102 (2020).

359. Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology* **23**, 258 (2022).

360. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

361. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

362. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

363. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Research* **42**, D222 (2014).

364. Haft, D. H. *et al.* TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res* **29**, 41–43 (2001).

365. Seemann, T. Snippy: fast bacterial variant calling from NGS reads. (2015).

366. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research* **43**, e15 (2015).

367. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom* **2**, e000056 (2016).

368. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

369. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).

370. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293–W296 (2021).

371. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).

372. Hadfield, J. *et al.* Phandango: an interactive viewer for bacterial population genomics. *Bioinformatics* **34**, 292–293 (2018).

373. Cochard, T., Branger, M., Supply, P., Sreevatsan, S. & Biet, F. MAC-INMV-SSR: a web application dedicated to genotyping members of *Mycobacterium avium* complex (MAC) including *Mycobacterium avium* subsp. *paratuberculosis* strains. *Infect Genet Evol* **77**, 104075 (2020).

374. Hunter, P. R. & Gaston, M. A. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *Journal of Clinical Microbiology* **26**, 2465–2466 (1988).

375. National Center for Biotechnology Information. SRA Toolkit. (2022).

376. Ozer, E. A. IN_SILICO_PCR. (2022).

377. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20-25 (2004).

378. Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, D265–D268 (2020).

379. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).

380. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

381. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

382. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026–1028 (2017).

383. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14–25 (2018).

384. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70–82 (2021).

385. Wayne, L. *et al.* Bergey's manual of systematic bacteriology. *Bergey's Manual of Systematic Bacteriology* **2**, (1986).

386. Sia, I. G. & Wieland, M. L. Current concepts in the management of tuberculosis. *Mayo Clin Proc* **86**, 348–361 (2011).

387. Jong, B. C. de, Antonio, M. & Gagneux, S. *Mycobacterium africanum*—Review of an Important Cause of Human Tuberculosis in West Africa. *PLOS Neglected Tropical Diseases* **4**, e744 (2010).

388. Cosivi, O. *et al.* Zoonotic tuberculosis due to *Mycobacterium bovis* in developing countries. *Emerg Infect Dis* **4**, 59–70 (1998).

389. Grange, J. M. *Mycobacterium bovis* infection in human beings. *Tuberculosis (Edinb)* **81**, 71–77 (2001).

390. Morris, R. S., Pfeiffer, D. U. & Jackson, R. The epidemiology of *Mycobacterium bovis* infections. *Vet Microbiol* **40**, 153–177 (1994).

391. Roe, W. D. *et al.* Pathology and molecular epidemiology of *Mycobacterium pinnipedii* tuberculosis in native New Zealand marine mammals. *PLoS One* **14**, e0212363 (2019).

392. Global Tuberculosis Programme. *Global tuberculosis report 2022*. 68 https://www.who.int/publications/i/item/9789240061729 (2022).

393. Tortoli, E. Microbiological Features and Clinical Relevance of New Species of the Genus *Mycobacterium*. *Clinical Microbiology Reviews* **27**, 727–752 (2014).

394. Reed, C. *et al.* Environmental Risk Factors for Infection with *Mycobacterium avium* Complex. *American Journal of Epidemiology* **164**, 32–40 (2006).

395. von Reyn, C. F. *et al.* Isolation of *Mycobacterium avium* complex from water in the United States, Finland, Zaire, and Kenya. *Journal of Clinical Microbiology* **31**, 3227–3230 (1993).

396. Griffith, D. E. *et al.* An Official ATS/IDSA Statement: Diagnosis, Treatment, and Prevention of Nontuberculous Mycobacterial Diseases. *Am J Respir Crit Care Med* **175**, 367–416 (2007).

397. Havlik, J. A., Jr. *et al.* Disseminated *Mycobacterium avium* Complex Infection: Clinical Identification and Epidemiologic Trends. *The Journal of Infectious Diseases* **165**, 577–580 (1992).

398. Jacobson, M. A. *et al.* Natural History of Disseminated *Mycobacterium avium* Complex Infection in AIDS. *The Journal of Infectious Diseases* **164**, 994–998 (1991).

399. Adekambi, T., Sassi, M., van Ingen, J. & Drancourt, M. Reinstating *Mycobacterium massiliense* and *Mycobacterium bolletii* as species of the *Mycobacterium abscessus* complex. *Int J Syst Evol Microbiol* **67**, 2726–2730 (2017).

400. Cho, Y.-J. *et al.* The Genome Sequence of '*Mycobacterium massiliense*' Strain CIP 108297 Suggests the Independent Taxonomic Status of the *Mycobacterium abscessus* Complex at the Subspecies Level. *PLOS ONE* **8**, e81560 (2013).

401. Lee, M.-R. *et al. Mycobacterium abscessus* Complex Infections in Humans. *Emerg Infect Dis* **21**, 1638–1646 (2015).

402. Sassi, M. & Drancourt, M. Genome analysis reveals three genomospecies in *Mycobacterium abscessus*. *BMC Genomics* **15**, 359 (2014).

403. Böttger, E. C. *Mycobacterium genavense*: an emerging pathogen. *Eur J Clin Microbiol Infect Dis* **13**, 932–936 (1994).

404. Hoefsloot, W. *et al. Mycobacterium genavense* in the Netherlands: an opportunistic pathogen in HIV and non-HIV immunocompromised patients. An observational study in 14 cases. *Clin Microbiol Infect* **19**, 432–437 (2013).

405. Hoop, R. K., Böttger, E. C., Ossent, P. & Salfinger, M. Mycobacteriosis due to *Mycobacterium genavense* in six pet birds. *J Clin Microbiol* **31**, 990–993 (1993).

406. Hughes, M. S. *et al.* Disseminated *Mycobacterium genavense* infection in a FIV-positive cat. *J Feline Med Surg* **1**, 23–29 (1999).

407. Kiehn, T. E. *et al. Mycobacterium genavense* infections in pet animals. *J Clin Microbiol* **34**, 1840–1842 (1996).

408. Krebs, T., Zimmerli, S., Bodmer, T. & Lämmle, B. *Mycobacterium genavense* infection in a patient with long-standing chronic lymphocytic leukaemia. *J Intern Med* **248**, 343–348 (2000).

409. Lucas, J. *et al. Mycobacterium genavense* infection in two aged ferrets with conjunctival lesions. *Aust Vet J* **78**, 685–689 (2000).

410. Atkins, B. L. & Gottlieb, T. Skin and soft tissue infections caused by nontuberculous mycobacteria. *Current Opinion in Infectious Diseases* **27**, 137 (2014).

411. Biet, F. & Boschiroli, M. L. Non-tuberculous mycobacterial infections of veterinary relevance. *Research in Veterinary Science* **97**, S69–S77 (2014).

412. Cook, J. L. Nontuberculous mycobacteria: opportunistic environmental pathogens for predisposed hosts. *British Medical Bulletin* **96**, 45–59 (2010).

413. Griffith, D. E. Nontuberculous mycobacterial lung disease. *Current Opinion in Infectious Diseases* **23**, 185 (2010).

414. Piersimoni, C. & Scarparo, C. Pulmonary infections associated with non-tuberculous mycobacteria in immunocompetent patients. *The Lancet Infectious Diseases* **8**, 323–334 (2008).

415. Coros, A., DeConno, E. & Derbyshire, K. M. IS6110, a *Mycobacterium tuberculosis* Complex-Specific Insertion Sequence, Is Also Present in the Genome of *Mycobacterium smegmatis*, Suggestive of Lateral Gene Transfer among Mycobacterial Species. *Journal of Bacteriology* **190**, 3408–3410 (2008).

416. Gonzalo-Asensio, J. *et al.* New insights into the transposition mechanisms of IS6110 and its dynamic distribution between *Mycobacterium tuberculosis* Complex lineages. *PLOS Genetics* **14**, e1007282 (2018).

417. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* **5**, 4812 (2014).

418. Guerrero, C., Bernasconi, C., Burki, D., Bodmer, T. & Telenti, A. A novel insertion element from *Mycobacterium avium*, *IS1245*, is a specific target for analysis of strain relatedness. *Journal of Clinical Microbiology* **33**, 304–307 (1995).

419. Biffa, D. *et al.* Multi-locus variable-number tandem repeat analysis (MLVA) reveals heterogeneity of *Mycobacterium bovis* strains and multiple genotype infections of cattle in Ethiopia. *Infection, Genetics and Evolution* **23**, 13–19 (2014).

420. Hill, V. *et al.* MLVA Based Classification of *Mycobacterium tuberculosis* Complex Lineages for a Robust Phylogeographic Snapshot of Its Worldwide Molecular Diversity. *PLOS ONE* **7**, e41991 (2012).

421. Overduin, P. *et al.* Use of Multilocus Variable-Number Tandem-Repeat Analysis for Typing *Mycobacterium avium* subsp. *paratuberculosis*. *Journal of Clinical Microbiology* **42**, 5022–5028 (2004).

422. Supply, P. *et al.* Automated High-Throughput Genotyping for Study of Global Epidemiology of *Mycobacterium tuberculosis* Based on Mycobacterial Interspersed Repetitive Units. *Journal of Clinical Microbiology* **39**, 3563–3571 (2001).

423. Supply, P. *et al.* Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* **44**, 4498–4510 (2006).

424. Homolka, S. *et al.* High Resolution Discrimination of Clinical *Mycobacterium tuberculosis* Complex Strains Based on Single Nucleotide Polymorphisms. *PLOS ONE* **7**, e39855 (2012).

425. Lipworth, S. *et al.* SNP-IT Tool for Identifying Subspecies and Associated Lineages of *Mycobacterium tuberculosis* Complex. *Emerg Infect Dis* **25**, 482–488 (2019).

426. Witney, A. A. *et al.* Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial. *BMC Medicine* **15**, 71 (2017).

427. Sobkowiak, B. *et al.* Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. *BMC Genomics* **19**, 613 (2018).

428. Fujita, K. *et al.* Association between Polyclonal and Mixed Mycobacterial *Mycobacterium avium* Complex Infection and Environmental Exposure. *Annals ATS* **11**, 45–53 (2014).

429. Jones, W. D. & Greenberg, J. Modification of methods used in bacteriophage typing of *Mycobacterium tuberculosis* isolates. *Journal of Clinical Microbiology* **7**, 467–469 (1978).

430. Bates, J. H., Stead, W. W. & Rado, T. A. Phage Type of Tubercle Bacilli Isolated from Patients with Two or More Sites of Organ Involvement. *Am Rev Respir Dis* **114**, 353–358 (1976).

431. Jones, W. D. Bacteriophage typing of *Mycobacterium tuberculosis* cultures from incidents of suspected laboratory cross-contamination. *Tubercle* **69**, 43–46 (1988).

432. Mankiewicz, E. & Liivak, M. Phage Types of *Mycobacterium tuberculosis* in Cultures Isolated from Eskimo Patients. *Am Rev Respir Dis* **111**, 307–312 (1975).

433. Snider, D. E., Jones, W. D. & Good, R. C. The Usefulness of Phage Typing *Mycobacterium tuberculosis* Isolates. *Am Rev Respir Dis* **130**, 1095–1099 (1984).

434. van Soolingen, D., de Haas, P. E., Hermans, P. W., Groenen, P. M. & van Embden, J. D. Comparison of various repetitive DNA elements as genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* **31**, 1987–1995 (1993).

435. van Soolingen, D. *et al.* Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. *Journal of Clinical Microbiology* **33**, 3234–3238 (1995).

436. Hermans, P. W. *et al.* Insertion element *IS986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *Journal of Clinical Microbiology* **28**, 2051–2058 (1990).

437. Thierry, D. *et al.* Characterization of a *Mycobacterium tuberculosis* insertion sequence, *IS6110*, and its application in diagnosis. *Journal of Clinical Microbiology* **28**, 2668–2673 (1990).

438. Thierry, D. *et al. IS6110*, an IS-like element of *Mycobacterium tuberculosis* complex. *Nucleic Acids Res* **18**, 188 (1990).

439. Brosch, R. Genomics, biology, and evolution of the *Mycobacterium tuberculosis* complex. *Molecular Genetics of Mycobacteria* 19–36 (2000).

440. Cave, M. D., Eisenach, K. D., McDermott, P. F., Bates, J. H. & Crawford, J. T. *IS6110*: Conservation of sequence in the *Mycobacterium tuberculosis* complex and its utilization in DNA fingerprinting. *Molecular and Cellular Probes* **5**, 73–80 (1991).

441. Cave, M. D. *et al.* Stability of DNA fingerprint pattern produced with *IS6110* in strains of *Mycobacterium tuberculosis*. *Journal of Clinical Microbiology* **32**, 262–266 (1994).

442. Fomukong, N. G. *et al.* Insertion sequence typing of *Mycobacterium tuberculosis*: characterization of a widespread subtype with a single copy of *IS6110*. *Tubercle and Lung Disease* **75**, 435–440 (1994).

443. Lok, K. H. *et al.* Molecular Differentiation of *Mycobacterium tuberculosis* Strains without *IS6110* Insertions. *Emerg Infect Dis* **8**, 1310–1313 (2002).

444. Singh, S. K., Verma, R. & Shah, D. H. Molecular fingerprinting of clinical isolates of *Mycobacterium bovis* and *Mycobacterium tuberculosis* from India by restriction fragment length polymorphism (RFLP). *J Vet Sci* **5**, 331–335 (2019).

445. Steensels, D., Fauville-Dufaux, M., Boie, J. & De Beenhouwer, H. Failure of PCR-Based *IS6110* Analysis To Detect Vertebral Spondylodiscitis Caused by *Mycobacterium bovis*. *Journal of Clinical Microbiology* **51**, 366–368 (2013).

446. van Soolingen, D., Hermans, P. W., de Haas, P. E., Soll, D. R. & van Embden, J. D. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *Journal of Clinical Microbiology* **29**, 2578–2586 (1991).

447. Yuen, L. K., Ross, B. C., Jackson, K. M. & Dwyer, B. Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. *Journal of Clinical Microbiology* **31**, 1615–1618 (1993).

448. van Embden, J. D. *et al.* Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *Journal of Clinical Microbiology* **31**, 406–409 (1993).

449. Allix, C., Supply, P. & Fauville-Dufaux, M. Utility of Fast Mycobacterial Interspersed Repetitive Unit—Variable Number Tandem Repeat Genotyping in Clinical Mycobacteriological Analysis. *Clinical Infectious Diseases* **39**, 783–789 (2004).

450. Almeida, S. M. de, Malaspina, A. C., Leite, C. Q. F. & Saad, M. H. F. Usefulness of 3'- 5' *IS6110*-RFLP genotyping and spoligotyping of *Mycobacterium tuberculosis* isolated in a tertiary hospital: a retrospective study detecting unsuspected epidemiological events. *Rev. Inst. Med. trop. S. Paulo* **61**, e51 (2019).

451. Andrade, M. K. N., Machado, S. M. A., Leite, M. L. & Saad, M. H. F. Phenotypic and genotypic variant of MDR-*Mycobacterium tuberculosis* multiple isolates in the same tuberculosis episode, Rio de Janeiro, Brazil. *Braz J Med Biol Res* **42**, 433–437 (2009).

452. Braden, C. R. *et al.* Simultaneous Infection with Multiple Strains of *Mycobacterium tuberculosis*. *Clinical Infectious Diseases* **33**, e42–e47 (2001).

453. Cerezo, I. *et al.* A first insight on the population structure of *Mycobacterium tuberculosis* complex as studied by spoligotyping and MIRU-VNTRs in Bogotá, Colombia. *Infection, Genetics and Evolution* **12**, 657–663 (2012).

454. Chaves, F., Dronda, F., Alonso-Sanz, M. & Noriega, A. R. Evidence of exogenous reinfection and mixed infection with more than one strain of *Mycobacterium tuberculosis* among Spanish HIV-infected inmates. *AIDS* **13**, 615 (1999).

455. Cox, H. S. *et al.* The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia. *Respir Res* **6**, 134 (2005).

456. Cox, H. S. *et al.* Risk of Acquired Drug Resistance during Short-Course Directly Observed Treatment of Tuberculosis in an Area with High Levels of Drug Resistance. *Clinical Infectious Diseases* **44**, 1421–1427 (2007).

457. Das, S. *et al.* Simultaneous infection with multiple strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis. *The International Journal of Tuberculosis and Lung Disease* **8**, 267–270 (2004).

458. du Plessis, D. G., Warren, R., Richardson, M., Joubert, J. J. & van Helden, P. D. Demonstration of reinfection and reactivation in HIV-negative autopsied cases of secondary tuberculosis: multilesional genotyping of *Mycobacterium tuberculosis* utilizing *IS 6110* and other repetitive element-based DNA fingerprinting. *Tuberculosis* **81**, 211–220 (2001).

459. García de Viedma, D., Marín, M., Ruiz Serrano, M. J., Alcalá, L. & Bouza, E. Polyclonal and Compartmentalized Infection by *Mycobacterium tuberculosis* in Patients with Both Respiratory and Extrarespiratory Involvement. *The Journal of Infectious Diseases* **187**, 695–699 (2003).

460. García de Viedma, D., Alonso Rodriguez, N., Andrés, S., Ruiz Serrano, M. J. & Bouza, E. Characterization of Clonal Complexity in Tuberculosis by Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing. *Journal of Clinical Microbiology* **43**, 5660–5664 (2005).

461. Gardy, J. L. *et al.* Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. *New England Journal of Medicine* **364**, 730–739 (2011).

462. Huyen, M. N. T. *et al.* Mixed Tuberculosis Infections in Rural South Vietnam. *Journal of Clinical Microbiology* **50**, 1586–1592 (2012).

463. Mokrousov, I. *et al.* Penitentiary population of *Mycobacterium tuberculosis* in Kyrgyzstan: Exceptionally high prevalence of the Beijing genotype and its Russia-specific subtype. *Infection, Genetics and Evolution* **9**, 1400–1405 (2009).

464. Navarro, Y. *et al.* Systematic Survey of Clonal Complexity in Tuberculosis at a Populational Level and Detailed Characterization of the Isolates Involved. *Journal of Clinical Microbiology* **49**, 4131–4137 (2011).

465. Pavlic, M., Allerberger, F., Dierich, M. P. & Prodinger, W. M. Simultaneous Infection with Two Drug-Susceptible *Mycobacterium tuberculosis* Strains in an Immunocompetent Host. *Journal of Clinical Microbiology* **37**, 4156–4157 (1999).

466. Richardson, M. *et al.* Multiple *Mycobacterium tuberculosis* Strains in Early Cultures from Patients in a High-Incidence Community Setting. *Journal of Clinical Microbiology* **40**, 2750–2754 (2002).

467. Shamputa, I. C. *et al.* Genotypic and Phenotypic Heterogeneity among *Mycobacterium tuberculosis* Isolates from Pulmonary Tuberculosis Patients. *Journal of Clinical Microbiology* **42**, 5528–5536 (2004).

468. Shamputa, I. C. *et al.* Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. *Respiratory Research* **7**, 99 (2006).

469. van der Zanden, A. G. M. *et al.* Multiple *Mycobacterium tuberculosis* infections in an HIV-infected patient. *Southeast Asian J Trop Med Public Health* **38**, 704–705 (2007).

470. von Reyn, C. F. *et al.* Disseminated tuberculosis in human immunodeficiency virus infection: ineffective immunity, polyclonal disease and high mortality. *The International Journal of Tuberculosis and Lung Disease* **15**, 1087–1092 (2011).

471. Yeh, R. W., Hopewell, P. C. & Daley, C. L. Simultaneous infection with two strains of *Mycobacterium tuberculosis* identified by restriction fragment length

polymorphism analysis. *The International Journal of Tuberculosis and Lung Disease* **3**, 537–539 (1999).

472. Mokrousov, I. *et al.* PCR-Based Methodology for Detecting Multidrug-Resistant Strains of *Mycobacterium tuberculosis* Beijing Family Circulating in Russia. *Eur J Clin Microbiol Infect Dis* **22**, 342–348 (2003).

473. Hofmann-Thiel, S. *et al.* Mechanisms of heteroresistance to isoniazid and rifampin of *Mycobacterium tuberculosis* in Tashkent, Uzbekistan. *European Respiratory Journal* **33**, 368–374 (2009).

474. Kamerbeek, J. *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *Journal of Clinical Microbiology* **35**, 907–914 (1997).

475. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLOS Computational Biology* **1**, e60 (2005).

476. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).

477. Hermans, P. W. *et al.* Insertion element *IS987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infection and Immunity* **59**, 2695–2705 (1991).

478. Andrews, J. R. *et al.* Exogenous Reinfection as a Cause of Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis in Rural South Africa. *The Journal of Infectious Diseases* **198**, 1582–1589 (2008).

479. Small, P. M. *et al.* The Epidemiology of Tuberculosis in San Francisco – A Population-Based Study Using Conventional and Molecular Methods. *New England Journal of Medicine* **330**, 1703–1709 (1994).

480. Warren, R. M. *et al.* Use of Spoligotyping for Accurate Classification of Recurrent Tuberculosis. *Journal of Clinical Microbiology* **40**, 3851–3853 (2002).

481. Caminero, J. A. *et al.* Epidemiological Evidence of the Spread of a *Mycobacterium tuberculosis* Strain of the Beijing Genotype on Gran Canaria Island. *Am J Respir Crit Care Med* **164**, 1165–1170 (2001).

482. Goyal, M. *et al.* Spoligotyping in molecular epidemiology of tuberculosis in Ghana. *Journal of Infection* **38**, 171–175 (1999).

483. Källenius, G. *et al.* Evolution and Clonal Traits of *Mycobacterium tuberculosis* Complex in Guinea-Bissau. *Journal of Clinical Microbiology* **37**, 3872–3878 (1999).

484. Nivin, B., Driscoll, J., Glaser, T., Bifani, P. & Munsiff, S. Use of Spoligotype Analysis to Detect Laboratory Cross-Contamination. *Infection Control & Hospital Epidemiology* **21**, 525–527 (2000).

485. Baffoe-Bonnie, A., Houpt, E. R., Turner, L., Dodge, D. & Heysell, S. K. Drug-Susceptible and Multidrug-Resistant *Mycobacterium tuberculosis* in a Single Patient. *Emerg Infect Dis* **25**, 2120–2121 (2019).

486. Chaoui, I. *et al.* Contribution of spoligotyping and MIRU-VNTRs to characterize prevalent *Mycobacterium tuberculosis* genotypes infecting tuberculosis patients in Morocco. *Infection, Genetics and Evolution* **21**, 463–471 (2014).

487. Egbe, N. F. *et al.* Molecular epidemiology of *Mycobacterium bovis* in Cameroon. *Sci Rep* **7**, 4652 (2017).

488. Furphy, C., Costello, E., Murphy, D., Corner, L. A. L. & Gormley, E. DNA Typing of *Mycobacterium bovis* Isolates from Badgers (*Meles meles*) Culled from Areas in Ireland with Different Levels of Tuberculosis Prevalence. *Veterinary Medicine International* **2012**, e742478 (2012).

489. Lamine-Khemiri, H. *et al.* Genotypic characterization by spoligotyping and VNTR typing of *Mycobacterium bovis* and *Mycobacterium caprae* isolates from cattle of Tunisia. *Trop Anim Health Prod* **46**, 305–311 (2014).

490. Ssengooba, W. *et al.* High Genotypic Discordance of Concurrent *Mycobacterium tuberculosis* Isolates from Sputum and Blood of HIV-Infected Individuals. *PLOS ONE* **10**, e0132581 (2015).

491. Umubyeyi, A. N. *et al.* Molecular investigation of recurrent tuberculosis in patients from Rwanda. *The International Journal of Tuberculosis and Lung Disease* **11**, 860–867 (2007).

492. Wang, X. *et al.* An investigation on the population structure of mixed infections of *Mycobacterium tuberculosis* in Inner Mongolia, China. *Tuberculosis* **95**, 695–700 (2015).

493. Navarro, Y. *et al.* Multiple sampling and discriminatory fingerprinting reveals clonally complex and compartmentalized infections by *M. bovis* in cattle. *Veterinary Microbiology* **175**, 99–104 (2015).

494. Silva-Pereira, T. T. *et al.* Genome sequencing of *Mycobacterium pinnipedii* strains: genetic characterization and evidence of superinfection in a South American sea lion (*Otaria flavescens*). *BMC Genomics* **20**, 1030 (2019).

495. Guernier-Cambert, V. *et al.* Diversity of *Mycobacterium tuberculosis* in the Middle Fly District of Western Province, Papua New Guinea: microbead-based spoligotyping using DNA from Ziehl-Neelsen-stained microscopy preparations. *Sci Rep* **9**, 15549 (2019).

496. Kargarpour Kamakoli, M. *et al.* Challenge in direct Spoligotyping of *Mycobacterium tuberculosis*: a problematic issue in the region with high prevalence of polyclonal infections. *BMC Research Notes* **11**, 486 (2018).

497. Lazzarini, L. C. O. *et al. Mycobacterium tuberculosis* spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. *Infection, Genetics and Evolution* **12**, 798–806 (2012).

498. Shamputa, I. C. *et al.* Genetic Diversity of *Mycobacterium tuberculosis* Isolates from a Tertiary Care Tuberculosis Hospital in South Korea. *Journal of Clinical Microbiology* **48**, 387–394 (2010).

499. Warren, R. M. *et al.* Patients with Active Tuberculosis often Have Different Strains in the Same Sputum Specimen. *Am J Respir Crit Care Med* **169**, 610–614 (2004).

500. Cox, R. & Mirkin, S. M. Characteristic enrichment of DNA repeats in different genomes. *Proceedings of the National Academy of Sciences* **94**, 5237–5242 (1997).

501. Nakamura, Y. *et al.* Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping. *Science* **235**, 1616–1622 (1987).

502. Supply, P., Magdalena, J., Himpens, S. & Locht, C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. *Molecular Microbiology* **26**, 991–1003 (1997).

503. Frothingham, R. & Meeker-O'Connell, W. A. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. *Microbiology* **144**, 1189–1196 (1998).

504. Le Flèche, P., Fabre, M., Denoeud, F., Koeck, J.-L. & Vergnaud, G. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiol* **2**, 37 (2002).

505. Roring, S. *et al.* Development of Variable-Number Tandem Repeat Typing of *Mycobacterium bovis*: Comparison of Results with Those Obtained by Using Existing Exact Tandem Repeats and Spoligotyping. *Journal of Clinical Microbiology* **40**, 2126–2133 (2002).

506. Skuce, R. A. *et al.* Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. *Microbiology* **148**, 519–528 (2002).

507. Smittipat, N. & Palittapongarnpim, P. Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis*. *Tubercle and Lung Disease* **80**, 69–74 (2000).

508. Abascal, E. *et al.* Screening of inmates transferred to Spain reveals a Peruvian prison as a reservoir of persistent *Mycobacterium tuberculosis* MDR strains and mixed infections. *Sci Rep* **10**, 2704 (2020).

509. Antusheva, E. *et al.* Outbreak of tuberculosis in a closed setting: views on transmission based on results from molecular and conventional methods. *Journal of Hospital Infection* **93**, 187–190 (2016).

510. Barletta, F. *et al.* Predominant *Mycobacterium tuberculosis* Families and High Rates of Recent Transmission among New Cases Are Not Associated with Primary Multidrug Resistance in Lima, Peru. *Journal of Clinical Microbiology* **53**, 1854–1863 (2015).

511. Cohen, T., Wilson, D., Wallengren, K., Samuel, E. Y. & Murray, M. Mixed-Strain *Mycobacterium tuberculosis* Infections among Patients Dying in a Hospital in KwaZulu-Natal, South Africa. *Journal of Clinical Microbiology* **49**, 385–388 (2011).

512. Dickman, K. R. *et al.* Detection of multiple strains of *Mycobacterium tuberculosis* using MIRU-VNTR in patients with pulmonary tuberculosis in Kampala, Uganda. *BMC Infectious Diseases* **10**, 349 (2010).

513. Ghielmetti, G. *et al.* Tuberculosis in Swiss captive Asian elephants: microevolution of *Mycobacterium tuberculosis* characterized by multilocus variable-number tandem-repeat analysis and whole-genome sequencing. *Sci Rep* **7**, 14647 (2017).

514. Hajimiri, E. S. *et al.* High prevalence of *Mycobacterium tuberculosis* mixed infection in the capital of moderate tuberculosis incidence country. *Microbial Pathogenesis* **93**, 213–218 (2016).

515. Hingley-Wilson, S. M. *et al.* Undetected Multidrug-Resistant Tuberculosis Amplified by First-line Therapy in Mixed Infection. *Emerg Infect Dis* **19**, 1138–1141 (2013).

516. Hu, Y. *et al.* Drug resistance characteristics and cluster analysis of *M. tuberculosis* in Chinese patients with multiple episodes of anti-tuberculosis treatment. *BMC Infectious Diseases* **16**, 4 (2016).

517. Kargarpour Kamakoli, M. *et al.* Comparison of MIRU-VNTR genotyping between old and fresh clinical samples in tuberculosis. *Infectious Diseases* **51**, 659–667 (2019).

518. Kargarpour Kamakoli, M. *et al.* Prediction of the hidden genotype of mixed infection strains in Iranian tuberculosis patients. *International Journal of Infectious Diseases* **95**, 22–27 (2020).

519. Khosravi, A. D. *et al.* Genetic diversity of multidrug-resistant *Mycobacterium tuberculosis* strains isolated from tuberculosis patients in Iran using MIRU-VNTR technique. *The Kaohsiung Journal of Medical Sciences* **33**, 550–557 (2017).

520. Kontsevaya, I. *et al.* Tuberculosis cases caused by heterogeneous infection in Eastern Europe and their influence on outcomes. *Infection, Genetics and Evolution* **48**, 76–82 (2017).

521. Mei, Z. *et al.* Discrepancies in Drug Susceptibility Test for Tuberculosis Patients Resulted from the Mixed Infection and the Testing System. *BioMed Research International* **2015**, e651980 (2015).

522. Mulenga, C. *et al.* Diversity of *Mycobacterium tuberculosis* genotypes circulating in Ndola, Zambia. *BMC Infect Dis* **10**, 177 (2010).

523. Muwonge, A. *et al.* Molecular investigation of multiple strain infections in patients with tuberculosis in Mubende district, Uganda. *Infection, Genetics and Evolution* **17**, 16–22 (2013).

524. Pang, Y. *et al.* Prevalence and risk factors of mixed *Mycobacterium tuberculosis* complex infections in China. *Journal of Infection* **71**, 231–237 (2015).

525. Peng, Y. *et al.* Multiple samples improve the sensitivity for detection of mixed *Mycobacterium* infections. *Tuberculosis* **93**, 548–550 (2013).

526. Sadegh, H. *et al.* Pros and cons of direct genotyping on tuberculosis clinical samples. *Microbial Pathogenesis* **103**, 135–138 (2017).

527. Shin, S. S. *et al.* Advanced Immune Suppression is Associated With Increased Prevalence of Mixed-Strain *Mycobacterium tuberculosis* Infections Among Persons at High Risk for Drug-Resistant Tuberculosis in Botswana. *The Journal of Infectious Diseases* **211**, 347–351 (2015).

528. Sichewo, P. R., Hlokwe, T. M., Etter, E. M. C. & Michel, A. L. Tracing cross species transmission of *Mycobacterium bovis* at the wildlife/livestock interface in South Africa. *BMC Microbiol* **20**, 49 (2020).

529. Stavrum, R. *et al.* High Diversity of *Mycobacterium tuberculosis* Genotypes in South Africa and Preponderance of Mixed Infections among ST53 Isolates. *Journal of Clinical Microbiology* **47**, 1848–1856 (2009).

530. Zetola, N. M. *et al.* Mixed *Mycobacterium tuberculosis* Complex Infections and False-Negative Results for Rifampin Resistance by GeneXpert MTB/RIF Are Associated with Poor Clinical Outcomes. *Journal of Clinical Microbiology* **52**, 2422–2429 (2014).

531. Zheng, C. *et al.* Mixed Infections and Rifampin Heteroresistance among *Mycobacterium tuberculosis* Clinical Isolates. *Journal of Clinical Microbiology* **53**, 2138–2147 (2015).

532. Mustafa, S. *et al.* Emergence of mixed infection of Beijing/Non-Beijing strains among multi-drug resistant *Mycobacterium tuberculosis* in Pakistan. *3 Biotech* **6**, 108 (2016).

533. Wang, J.-Y. *et al.* Mixed infection with Beijing and non-Beijing strains in pulmonary tuberculosis in Taiwan: prevalence, risk factors, and dominant strain. *Clinical Microbiology and Infection* **17**, 1239–1245 (2011).

534. Mallard, K. *et al.* Molecular Detection of Mixed Infections of *Mycobacterium tuberculosis* Strains in Sputum Samples from Patients in Karonga District, Malawi. *Journal of Clinical Microbiology* **48**, 4512–4518 (2010).

535. Hanekom, M. *et al.* Population Structure of Mixed *Mycobacterium tuberculosis* Infection Is Strain Genotype and Culture Medium Dependent. *PLOS ONE* **8**, e70178 (2013).

536. Chen, J. *et al.* Deletion-targeted multiplex PCR (DTM-PCR) for identification of Beijing/W genotypes of *Mycobacterium tuberculosis*. *Tuberculosis* **87**, 446–449 (2007).

537. Theisen, A. *et al.* Mixed-strain infection with a drug-sensitive and multidrug-resistant strain of *Mycobacterium tuberculosis*. *The Lancet* **345**, 1512–1513 (1995).

538. Müller, R., Roberts, C. A. & Brown, T. A. Genotyping of ancient *Mycobacterium tuberculosis* strains reveals historic genetic diversity. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20133236 (2014).

539. Sreevatsan, S. *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences* **94**, 9869–9874 (1997).

540. Lourenço, M. C. S. *et al.* Genotypic patterns of multiple isolates of *M. tuberculosis* from tuberculous HIV patients. *Tropical Medicine & International Health* **5**, 488–494 (2000).

541. Friedman, C. R., Stoeckle, M. Y., Johnson, W. D. & Riley, L. W. Double-repetitive-element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. *Journal of Clinical Microbiology* **33**, 1383–1384 (1995).

542. Hatherell, H.-A. *et al.* Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Medicine* **14**, 21 (2016).

543. Bryant, J. M. *et al.* Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *The Lancet Respiratory Medicine* **1**, 786–792 (2013).

544. Dippenaar, A. *et al.* Whole genome sequencing provides additional insights into recurrent tuberculosis classified as endogenous reactivation by IS6110 DNA fingerprinting. *Infection, Genetics and Evolution* **75**, 103948 (2019).

545. Guerra-Assunção, J. A. *et al.* Recurrence due to Relapse or Reinfection With *Mycobacterium tuberculosis*: A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort With a High HIV Infection Prevalence and Active Follow-up. *The Journal of Infectious Diseases* **211**, 1154–1163 (2015).

546. Gan, M., Liu, Q., Yang, C., Gao, Q. & Luo, T. Deep Whole-Genome Sequencing to Detect Mixed Infection of *Mycobacterium tuberculosis*. *PLOS ONE* **11**, e0159029 (2016).

547. Nimmo, C. *et al.* Dynamics of within-host *Mycobacterium tuberculosis* diversity and heteroresistance during treatment. *EBioMedicine* **55**, 102747 (2020).

548. Döpfer, D. *et al.* Assessing Genetic Heterogeneity within Bacterial Species Isolated from Gastrointestinal and Environmental Samples: How Many Isolates Does It Take? *Applied and Environmental Microbiology* **74**, 3490–3496 (2008).

549. van den Berg, R. J. *et al.* Coexistence of multiple PCR-ribotype strains of *Clostridium difficile* in faecal samples limits epidemiological studies. *Journal of Medical Microbiology* **54**, 173–179 (2005).

550. Anyansi, C. *et al.* QuantTB – a method to classify mixed *Mycobacterium tuberculosis* infections within whole genome sequencing data. *BMC Genomics* **21**, 80 (2020).

551. O'Donnell, M. R. *et al.* Early Detection of Emergent Extensively Drug-Resistant Tuberculosis by Flow Cytometry-Based Phenotyping and Whole-Genome Sequencing. *Antimicrobial Agents and Chemotherapy* **63**, e01834-18 (2019).

552. Wollenberg, K. *et al.* A retrospective genomic analysis of drug-resistant strains of *M. tuberculosis* in a high-burden setting, with an emphasis on comparative diagnostics and reactivation and reinfection status. *BMC Infect Dis* **20**, 17 (2020).

553. Pfeiffer, W., Braun, J., Burchell, J., Witte, C. L. & Rideout, B. A. Whole-genome analysis of mycobacteria from birds at the San Diego Zoo. *PLOS ONE* **12**, e0173464 (2017).

554. Chan, J. Z.-M. *et al.* Metagenomic Analysis of Tuberculosis in a Mummy. *New England Journal of Medicine* **369**, 289–290 (2013).

555. Dheda, K. *et al.* Outcomes, infectiousness, and transmission dynamics of patients with extensively drug-resistant tuberculosis and home-discharged patients with

programmatically incurable tuberculosis: a prospective cohort study. *The Lancet Respiratory Medicine* **5**, 269–281 (2017).

556. Kay, G. L. *et al.* Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. *Nat Commun* **6**, 6717 (2015).

557. Nicol, M. P., Cox, H., Nicol, M. P. & Cox, H. Recent developments in the diagnosis of drug-resistant tuberculosis. *Microbiol. Aust.* **40**, 82–86 (2019).

558. Arbeit, R. D. *et al.* Genetic Diversity among Strains of *Mycobacterium avium* Causing Monoclonal and Polyclonal Bacteremia in Patients with AIDS. *The Journal of Infectious Diseases* **167**, 1384–1390 (1993).

559. Devallois, A. & Rastogi, N. Computer-assisted analysis of *Mycobacterium avium* fingerprints using insertion elements *IS1245* and *IS1311* in a Caribbean setting. *Research in Microbiology* **148**, 703–713 (1997).

560. Dvorska, L. *et al. IS1311* and *IS1245* Restriction Fragment Length Polymorphism Analyses, Serotypes, and Drug Susceptibilities of *Mycobacterium avium* Complex Isolates Obtained from a Human Immunodeficiency Virus-Negative Patient. *Journal of Clinical Microbiology* **40**, 3712–3719 (2002).

561. García-Pedrazuela, M. *et al.* Polyclonality among clinical strains of non-pigmented rapidly growing mycobacteria: phenotypic and genotypic differences and their potential implications. *Clinical Microbiology and Infection* **21**, 348.e1-348.e4 (2015).

562. Kimizuka, Y. *et al.* Retrospective evaluation of natural course in mild cases of *Mycobacterium avium* complex pulmonary disease. *PLOS ONE* **14**, e0216034 (2019).

563. Legrand, E., Devallois, A., Horgen, L. & Rastogi, N. A Molecular Epidemiological Study of *Mycobacterium simiae* Isolated from AIDS Patients in Guadeloupe. *Journal of Clinical Microbiology* **38**, 3080–3084 (2000).

564. Legrand, E., Sola, C., Verdol, B. & Rastogi, N. Genetic diversity of *Mycobacterium avium* recovered from AIDS patients in the Caribbean as studied by a consensus *IS1245*-RFLP method and pulsed-field gel electrophoresis. *Research in Microbiology* **151**, 271–283 (2000).

565. Ohkusu, K., Bermudez, L. E., Nash, K. A., MacGregor, R. R. & Inderlied, C. B. Differential Virulence of *Mycobacterium avium* Strains Isolated from HIV-Infected Patients with Disseminated *M. avium* Complex Disease. *The Journal of Infectious Diseases* **190**, 1347–1354 (2004).

566. Oliveira, R. S. *et al.* PCR-Restriction Enzyme Analysis of a Bone Marrow Isolate from a Human Immunodeficiency Virus-Positive Patient Discloses Polyclonal Infection with Two *Mycobacterium avium* Strains. *Journal of Clinical Microbiology* **38**, 4643–4645 (2000).

567. Panunto, A. C., Villares, M. C. B. & Ramos, M. C. *IS1245* restriction fragment length polymorphism typing of *Mycobacterium avium* from patients admitted to a reference hospital in Campinas, Brazil. *Braz J Med Biol Res* **36**, 1397–1401 (2003).

568. Picardeau, M. *et al.* Use of different molecular typing techniques for bacteriological follow-up in a clinical trial with AIDS patients with *Mycobacterium avium* bacteremia. *Journal of Clinical Microbiology* **35**, 2503–2510 (1997).

569. Saad, M. H. F. *et al.* Multiple isolates from Aids patients: aspects of an analysis by a genotypic marker and antimicrobial susceptibilities variations. *Mem. Inst. Oswaldo Cruz* **95**, 729–732 (2000).

570. Sequeira, P. C. de, Fonseca, L. de S., Silva, M. G. da & Saad, M. H. F. *Mycobacterium avium* restriction fragment length polymorphism-IS *IS1245* and the simple double repetitive element polymerase chain reaction typing method to screen genetic diversity in Brazilian strains. *Mem. Inst. Oswaldo Cruz* **100**, 743–748 (2005).

571. Slutsky, A. M. *et al.* Polyclonal infections due to *Mycobacterium avium* complex in patients with AIDS detected by pulsed-field gel electrophoresis of sequential clinical isolates. *Journal of Clinical Microbiology* **32**, 1773–1778 (1994).

572. von Reyn, C. F., Jacobs, N. J., Arbeit, R. D., Maslow, J. N. & Niemczyk, S. Polyclonal *Mycobacterium avium* infections in patients with AIDS: variations in antimicrobial susceptibilities of different strains of *M. avium* isolated from the same patient. *Journal of Clinical Microbiology* **33**, 1008–1010 (1995).

573. Wallace, R. J. *et al.* Polyclonal *Mycobacterium avium* complex infections in patients with nodular bronchiectasis. *Am J Respir Crit Care Med* **158**, 1235–1244 (1998).

574. Dvorska, L. *et al.* Avian tuberculosis in naturally infected captive water birds of the Ardeideae and Threskiornithidae families studied by serotyping, *IS901* RFLP typing, and virulence for poultry. *Veterinary Microbiology* **119**, 366–374 (2007).

575. Gerritsmann, H. *et al.* Multiple strain infections and high genotypic diversity among *Mycobacterium avium* subsp. *paratuberculosis* field isolates from diseased wild and

domestic ruminant species in the eastern Alpine region of Austria. *Infection, Genetics and Evolution* **21**, 244–251 (2014).

576. Gioffré, A. *et al.* Molecular typing of Argentinian *Mycobacterium avium* subsp. *paratuberculosis* isolates by multiple-locus variable number-tandem repeat analysis. *Braz. J. Microbiol.* **46**, 557–564 (2015).

577. Johansen, T. B., Agdestein, A., Lium, B., Jørgensen, A. & Djønne, B. *Mycobacterium avium* subsp. *hominissuis* Infection in Swine Associated with Peat Used for Bedding. *BioMed Research International* **2014**, e189649 (2014).

578. Shitaye, J. E. *et al. Mycobacterium avium* subsp. *avium* distribution studied in a naturally infected hen flock and in the environment by culture, serotyping and *IS901* RFLP methods. *Veterinary Microbiology* **127**, 155–164 (2008).

579. Pate, M., Žolnir-Dovč, M., Krt, B. & Ocepek, M. *IS1245* RFLP-based genotyping study of *Mycobacterium avium* subsp. *hominissuis* isolates from pigs and humans. *Comparative Immunology, Microbiology and Infectious Diseases* **31**, 537–550 (2008).

580. Picardeau, M. & Vincent, V. Typing of *Mycobacterium avium* isolates by PCR. *Journal of Clinical Microbiology* **34**, 389–392 (1996).

581. Tenover, F. C. *et al.* Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**, 2233–2239 (1995).

582. Zhang, Y., Rajagopalan, M., Brown, B. A. & Wallace, R. J. Randomly amplified polymorphic DNA PCR for comparison of *Mycobacterium abscessus* strains from nosocomial outbreaks. *Journal of Clinical Microbiology* **35**, 3132–3139 (1997).

583. Oakey, J., Gavey, L., Singh, S. V., Platell, J. & Waltisbuhl, D. Variable-number tandem repeats genotyping used to aid and inform management strategies for a bovine Johne's disease incursion in tropical and subtropical Australia. *J Vet Diagn Invest* **26**, 651–657 (2014).

584. Walker, T. M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *The Lancet Infectious Diseases* **13**, 137–146 (2013).

585. Scollard, D. M. *et al.* The Continuing Challenges of Leprosy. *Clinical Microbiology Reviews* **19**, 338–381 (2006).

586. Marx, F. M. *et al.* The Temporal Dynamics of Relapse and Reinfection Tuberculosis After Successful Treatment: A Retrospective Cohort Study. *Clinical Infectious Diseases* **58**, 1676–1683 (2014).

587. Maghradze, N. *et al.* Classifying recurrent *Mycobacterium tuberculosis* cases in Georgia using MIRU-VNTR typing. *PLOS ONE* **14**, e0223610 (2019).

588. Ford, C. B. *et al. Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* **45**, 784–790 (2013).

589. Roetzer, A. *et al.* Whole Genome Sequencing versus Traditional Genotyping for Investigation of a *Mycobacterium tuberculosis* Outbreak: A Longitudinal Molecular Epidemiological Study. *PLOS Medicine* **10**, e1001387 (2013).

590. Jajou, R. *et al.* Occurrence and Nature of Double Alleles in Variable-Number Tandem-Repeat Patterns of More than 8,000 *Mycobacterium tuberculosis* Complex Isolates in The Netherlands. *Journal of Clinical Microbiology* **56**, e00761-17 (2018).

591. Martín, A., Herránz, M., Serrano, M. J. R., Bouza, E. & de Viedma, D. G. Rapid clonal analysis of recurrent tuberculosis by direct MIRU-VNTR typing on stored isolates. *BMC Microbiol* **7**, 73 (2007).

592. Al-Hajoj, S. A. M. *et al.* Microevolution of *Mycobacterium tuberculosis* in a Tuberculosis Patient. *Journal of Clinical Microbiology* **48**, 3813–3816 (2010).

593. Alves, S. L. de Á., Metzker, F. S., Araújo-Filho, J. A. de, Junqueira-Kipnis, A. P. & Kipnis, A. Clinical data and molecular analysis of *Mycobacterium tuberculosis* isolates from drug-resistant tuberculosis patients in Goiás, Brazil. *Mem. Inst. Oswaldo Cruz* **106**, 655–661 (2011).

594. Liu, Q. *et al.* Within patient microevolution of *Mycobacterium tuberculosis* correlates with heterogeneous responses to treatment. *Sci Rep* **5**, 17507 (2015).

595. Pérez-Lago, L., Navarro, Y., Herranz, M., Bouza, E. & García-de-Viedma, D. Differences in gene expression between clonal variants of *Mycobacterium tuberculosis* emerging as a result of microevolution. *International Journal of Medical Microbiology* **303**, 674–677 (2013).

596. Fonseca, J. D., Knight, G. M. & McHugh, T. D. The complex evolution of antibiotic resistance in *Mycobacterium tuberculosis*. *International Journal of Infectious Diseases* **32**, 94–100 (2015).

597. Pérez-Lago, L. *et al.* Whole Genome Sequencing Analysis of Intrapatient Microevolution in *Mycobacterium tuberculosis*: Potential Impact on the Inference of Tuberculosis Transmission. *The Journal of Infectious Diseases* **209**, 98–108 (2014).

598. Ssengooba, W., de Jong, B. C., Joloba, M. L., Cobelens, F. G. & Meehan, C. J. Whole genome sequencing reveals mycobacterial microevolution among concurrent

isolates from sputum and blood in HIV infected TB patients. *BMC Infectious Diseases* **16**, 371 (2016).

599. Buff, A. M. *et al.* Two Tuberculosis Genotyping Clusters, One Preventable Outbreak. *Public Health Rep* **124**, 490–494 (2009).

600. Moravkova, M., Mrlik, V., Parmova, I., Kriz, P. & Pavlik, I. High incidence of *Mycobacterium avium* subspecies *hominissuis* infection in a zoo population of bongo antelopes (Tragelaphus eurycerus). *J Vet Diagn Invest* **25**, 531–534 (2013).

601. Reis, A. C., Albuquerque, T., Botelho, A. & Cunha, M. V. Polyclonal infection as a new scenario in *Mycobacterium caprae* epidemiology. *Veterinary Microbiology* **240**, 108533 (2020).

602. Agrawal, G., Aitken, J., Hamblin, H., Collins, M. & Borody, T. J. Putting Crohn's on the MAP: Five Common Questions on the Contribution of *Mycobacterium avium* subspecies paratuberculosis to the Pathophysiology of Crohn's Disease. *Dig Dis Sci* **66**, 348–358 (2021).

603. Ekundayo, T. C. *et al.* Systematic and meta-analysis of *Mycobacterium avium* subsp. *paratuberculosis* related type 1 and type 2 diabetes mellitus. *Sci Rep* **12**, 4608 (2022).

604. Kuenstner, J. T. *et al.* The Consensus from the *Mycobacterium avium* ssp. *paratuberculosis* (MAP) Conference 2017. *Front Public Health* **5**, 208 (2017).

605. Waddell, L. A., Rajić, A., Stärk, K. D. C. & McEwen, S. A. The potential Public Health Impact of *Mycobacterium avium* ssp. *paratuberculosis*: Global Opinion Survey of Topic Specialists. *Zoonoses Public Health* **63**, 212–222 (2016).

606. Mitchell, R. M. *et al.* Elucidating Transmission Patterns of Endemic *Mycobacterium avium* subsp. *paratuberculosis* Using Molecular Epidemiology. *Veterinary Sciences* **6**, 32 (2019).

607. Jakobsen, M. B., Alban, L. & Nielsen, S. S. A cross-sectional study of paratuberculosis in 1155 Danish dairy cows. *Preventive Veterinary Medicine* **46**, 15–27 (2000).

608. Salgado, M., Kruze, J. & Collins, M. T. Diagnosis of paratuberculosis by fecal culture and ELISA on milk and serum samples in two types of Chilean dairy goat herds. *J Vet Diagn Invest* **19**, 99–102 (2007).

609. Kralik, P., Pribylova-Dziedzinska, R., Kralova, A., Kovarcik, K. & Slana, I. Evidence of passive faecal shedding of *Mycobacterium avium* subsp. *paratuberculosis* in a Limousin cattle herd. *The Veterinary Journal* **201**, 91–94 (2014).

610. Whittington, R. Cultivation of *Mycobacterium avium* subsp. *paratuberculosis*. *Paratuberculosis: organism, disease, control* 266–304 (2020) doi:10.1079/9781789243413.0266.

611. Whitlock, R., Rosenberger, A. & Spencer, P. Laboratory culture techniques for Johne's disease: a critical evaluation of contamination and incubation times. in *Proceedings-Annual Meeting of the United States Animal Health Association* vol. 93 382–386 (1989).

612. Whittington, R. J. Factors Affecting Isolation and Identification of *Mycobacterium avium* subsp. *paratuberculosis* from Fecal and Tissue Samples in a Liquid Culture System. *Journal of Clinical Microbiology* **47**, 614–622 (2009).

613. Khare, S., Adams, L. G., Osterstock, J., Roussel, A. & David, L. Effects of Shipping and Storage Conditions of Fecal Samples on Viability of *Mycobacterium paratuberculosis*. *J Clin Microbiol* **46**, 1561–1562 (2008).

614. Sohal, J. S., Singh, S. V., Singh, P. K. & Singh, A. V. On the evolution of 'Indian Bison type' strains of *Mycobacterium avium* subspecies *paratuberculosis*. *Microbiol Res* **165**, 163–171 (2010).

615. Mitchell, R. M. *et al.* Differences in intermittent and continuous fecal shedding patterns between natural and experimental *Mycobacterium avium* subspecies *paratuberculosis* infections in cattle. *Vet Res* **46**, 66 (2015).

616. Lim, J. *et al.* Genomic diversity of *Mycobacterium avium* subsp. *paratuberculosis*: pangenomic approach for highlighting unique genomic features with newly constructed complete genomes. *Veterinary Research* **52**, 46 (2021).

617. Sampson, S. L. Mycobacterial PE/PPE Proteins at the Host-Pathogen Interface. *Journal of Immunology Research* **2011**, e497203 (2011).

618. Timms, V. J., Hassan, K. A., Mitchell, H. M. & Neilan, B. A. Comparative genomics between human and animal associated subspecies of the *Mycobacterium avium* complex: a basis for pathogenicity. *BMC Genomics* **16**, 695 (2015).

619. Pal, R., Nazar, F. & Mukhopadhyay, S. The PE and PPE Family Proteins of *Mycobacterium tuberculosis*: What they Are Up To? in *Mycobacterium Tuberculosis: Molecular Infection Biology, Pathogenesis, Diagnostics and New Interventions* (eds. Hasnain, S. E., Ehtesham, N. Z. & Grover, S.) 123–150 (Springer, 2019). doi:10.1007/978-981-32-9413-4_8.

620. Bansal, K. *et al.* PE_PGRS Antigens of *Mycobacterium tuberculosis* Induce Maturation and Activation of Human Dendritic Cells. *The Journal of Immunology* **184**, 3495–3504 (2010).

621. Brennan, M. J. *et al.* Evidence that Mycobacterial PE_PGRS Proteins Are Cell Surface Constituents That Influence Interactions with Other Cells. *Infection and Immunity* **69**, 7326–7333 (2001).

622. Cossu, A., Sechi, L. A., Zanetti, S. & Rosu, V. Gene expression profiling of *Mycobacterium avium* subsp. *paratuberculosis* in simulated multi-stress conditions and within THP-1 cells reveals a new kind of interactive intramacrophage behaviour. *BMC Microbiology* **12**, 87 (2012).

623. Mackenzie, N., Alexander, D. C., Turenne, C. Y., Behr, M. A. & De Buck, J. M. Genomic Comparison of PE and PPE Genes in the *Mycobacterium avium* Complex. *Journal of Clinical Microbiology* **47**, 1002–1011 (2009).

624. Casali, N. & Riley, L. W. A phylogenomic analysis of the Actinomycetales mce operons. *BMC Genomics* **8**, 60 (2007).

625. Hemati, Z. *et al.* Mammalian cell entry operons; novel and major subset candidates for diagnostics with special reference to *Mycobacterium avium* subspecies *paratuberculosis* infection. *Veterinary Quarterly* **39**, 65–75 (2019).

626. Melly, G. & Purdy, G. E. MmpL Proteins in Physiology and Pathogenesis of *M. tuberculosis*. *Microorganisms* **7**, 70 (2019).

627. Ahlstrom, C., Barkema, H. W. & De Buck, J. Improved Short-Sequence-Repeat Genotyping of *Mycobacterium avium* subsp. *paratuberculosis* by Using Matrix-

Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry. *Applied and Environmental Microbiology* **80**, 534–539 (2014).

628. de Kruijf, M. *et al.* Low genetic diversity of bovine *Mycobacterium avium* subspecies *paratuberculosis* isolates detected by MIRU-VNTR genotyping. *Vet Microbiol* **203**, 280–285 (2017).
629. Kasnitz, N., Köhler, H., Weigoldt, M., Gerlach, G. F. & Möbius, P. Stability of genotyping target sequences of *Mycobacterium avium* subsp. *paratuberculosis* upon cultivation on different media, in vitro- and in vivo passage, and natural infection. *Veterinary Microbiology* **167**, 573–583 (2013).
630. Ahlstrom, C. *et al.* Limitations of variable number of tandem repeat typing identified through whole genome sequencing of *Mycobacterium avium* subsp. *paratuberculosis* on a national and herd level. *BMC Genomics* **16**, 161 (2015).
631. Marri, P. R., Bannantine, J. P. & Golding, G. B. Comparative genomics of metabolic pathways in *Mycobacterium* species: gene duplication, gene decay and lateral gene transfer. *FEMS Microbiol Rev* **30**, 906–925 (2006).
632. Eppleston, J., Begg, D. J., Dhand, N. K., Watt, B. & Whittington, R. J. Environmental Survival of *Mycobacterium avium* subsp. *paratuberculosis* in Different Climatic Zones of Eastern Australia. *Appl Environ Microbiol* **80**, 2337–2342 (2014).
633. Mizzi, R., Plain, K. M., Whittington, R. & Timms, V. J. Global Phylogeny of *Mycobacterium avium* and Identification of Mutation Hotspots During Niche Adaptation. *Frontiers in Microbiology* **13**, (2022).

634. Uchiya, K.-I. *et al.* Comparative genome analyses of *Mycobacterium avium* reveal genomic features of its subspecies and strains that cause progression of pulmonary disease. *Sci Rep* **7**, 39750 (2017).

635. Abukhalid, N., Islam, S., Ndzeidze, R. & Bermudez, L. E. *Mycobacterium avium* subsp. *hominissuis* Interactions with Macrophage Killing Mechanisms. *Pathogens* **10**, 1365 (2021).

636. Dayananda, P. & Wilcox, M. H. A Review of Mixed Strain *Clostridium difficile* Colonization and Infection. *Frontiers in Microbiology* **10**, (2019).

637. Raven, K. E., Gouliouris, T., Parkhill, J. & Peacock, S. J. Genome-Based Analysis of *Enterococcus faecium* Bacteremia Associated with Recurrent and Mixed-Strain Infection. *J Clin Microbiol* **56**, e01520-17 (2018).

638. Wynne, J. W. *et al.* Resequencing the *Mycobacterium avium* subsp. *paratuberculosis* K10 genome: improved annotation and revised genome sequence. *J Bacteriol* **192**, 6319–6320 (2010).

639. Argemi, X. *et al.* Comparative genomic analysis of *Staphylococcus lugdunensis* shows a closed pan-genome and multiple barriers to horizontal gene transfer. *BMC Genomics* **19**, 621 (2018).

640. McInerney, J. O., McNally, A. & O'Connell, M. J. Why prokaryotes have pangenomes. *Nat Microbiol* **2**, 1–5 (2017).

641. Saeed, A. F., Wang, R. & Wang, S. Microsatellites in Pursuit of Microbial Genome Evolution. *Frontiers in Microbiology* **6**, (2016).

642. Sreenu, V. B., Kumar, P., Nagaraju, J. & Nagarajam, H. A. Simple sequence repeats in mycobacterial genomes. *J Biosci* **32**, 3–15 (2007).

643. Arnold, C., Thorne, N., Underwood, A., Baster, K. & Gharbia, S. Evolution of short sequence repeats in *Mycobacterium tuberculosis*. *FEMS Microbiol Lett* **256**, 340–346 (2006).

644. Supply, P. *et al.* Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. *Mol Microbiol* **36**, 762–771 (2000).

645. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics and Bioinformatics* **3**, lqab019 (2021).

646. Modi, A., Vai, S., Caramelli, D. & Lari, M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. *Methods Mol Biol* **2242**, 15–42 (2021).

647. van Belkum, A. The role of short sequence repeats in epidemiologic typing. *Current Opinion in Microbiology* **2**, 306–311 (1999).

648. Hu, J. *et al.* Phenotypic diversity and correlation between white-opaque switching and the CAI microsatellite locus in *Candida albicans*. *Curr Genet* **62**, 585–593 (2016).

649. Lysnyansky, I., Rosengarten, R. & Yogev, D. Phenotypic switching of variable surface lipoproteins in *Mycoplasma bovis* involves high-frequency chromosomal rearrangements. *J Bacteriol* **178**, 5395–5401 (1996).

650. Henderson, I. R., Meehan, M. & Owen, P. A novel regulatory mechanism for a novel phase-variable outer membrane protein of *Escherichia coli*. *Adv Exp Med Biol* **412**, 349–355 (1997).

651. Torres-Cruz, J. & van der Woude, M. W. Slipped-Strand Mispairing Can Function as a Phase Variation Mechanism in *Escherichia coli*. *J Bacteriol* **185**, 6990–6994 (2003).

652. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420–423 (2019).

653. Hayashi, S. & Wu, H. C. Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, 451–471 (1990).

654. Kaushik, S., He, H. & Dalbey, R. E. Bacterial Signal Peptides- Navigating the Journey of Proteins. *Frontiers in Physiology* **13**, (2022).

655. Tjalsma, H. *et al.* The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of alpha-amylase, a non-lipoprotein. *J Biol Chem* **274**, 1698–1707 (1999).

656. Lorenz, K., Lohse, M. J. & Quitterer, U. Protein kinase C switches the Raf kinase inhibitor from Raf-1 to GRK-2. *Nature* **426**, 574–579 (2003).

657. Skinner, J. J. & Rosner, M. R. RKIP structure drives its function: a three-state model for regulation of RKIP. *Crit Rev Oncog* **19**, 483–488 (2014).

658. Yeung, K. C. *et al.* Raf kinase inhibitor protein interacts with NF-kappaB-inducing kinase and TAK1 and inhibits NF-kappaB activation. *Mol Cell Biol* **21**, 7207–7217 (2001).

659. Serre, L. *et al.* Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein. *J Mol Biol* **310**, 617–634 (2001).

660. Lau, M. E., Loughman, J. A. & Hunstad, D. A. YbcL of uropathogenic *Escherichia coli* suppresses transepithelial neutrophil migration. *Infect Immun* **80**, 4123–4132 (2012).

661. An, L.-P. *et al.* Purification, molecular cloning and functional characterization of swine phosphatidylethanolamine-binding protein 4 from seminal plasma. *Biochem Biophys Res Commun* **423**, 690–696 (2012).

662. He, H. *et al.* Phosphatidylethanolamine binding protein 4 (PEBP4) is a secreted protein and has multiple functions. *Biochim Biophys Acta* **1863**, 1682–1689 (2016).

663. Akdel, M. *et al.* A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* **29**, 1056–1067 (2022).

664. Ruff, K. M. & Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of Molecular Biology* **433**, 167208 (2021).

665. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

666. Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* **7**, 493–503 (2009).

667. van der Woude, M. W. & Bäumler, A. J. Phase and Antigenic Variation in Bacteria. *Clin Microbiol Rev* **17**, 581–611 (2004).

668. Gu, T., Tan, S., Gou, X., Araki, H. & Tian, D. Avoidance of Long Mononucleotide Repeats in Codon Pair Usage. *Genetics* **186**, 1077–1084 (2010).

669. Sagher, D., Hsu, A. & Strauss, B. Stabilization of the intermediate in frameshift mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **423**, 73–77 (1999).

670. Abo-Kadoum, M. A. *et al. Mycobacterium tuberculosis* Raf kinase inhibitor protein (RKIP) Rv2140c is involved in cell wall arabinogalactan biosynthesis via phosphorylation. *Microbiol Res* **242**, 126615 (2021).

671. Griffiths, E. C., Pedersen, A. B., Fenton, A. & Petchey, O. L. The nature and consequences of coinfection in humans. *J Infect* **63**, 200–206 (2011).

672. Kordick, S. K. *et al.* Coinfection with multiple tick-borne pathogens in a Walker Hound kennel in North Carolina. *J Clin Microbiol* **37**, 2631–2638 (1999).

673. Middlebrook, E. A. *et al.* Identification and distribution of pathogens coinfecting with *Brucella* spp., *Coxiella burnetii* and Rift Valley fever virus in humans, livestock and wildlife. *Zoonoses Public Health* **69**, 175–194 (2022).

674. Sabey, K. A., Song, S. J., Jolles, A., Knight, R. & Ezenwa, V. O. Coinfection and infection duration shape how pathogens affect the African buffalo gut microbiota. *ISME J* **15**, 1359–1371 (2021).

675. Hodgson, D. J. *et al.* Host ecology determines the relative fitness of virus genotypes in mixed-genotype nucleopolyhedrovirus infections. *J Evol Biol* **17**, 1018–1025 (2004).

676. McNaughton, A. L., Thomson, E. C., Templeton, K., Gunson, R. N. & Leitch, E. C. M. Mixed Genotype Hepatitis C Infections and Implications for Treatment. *Hepatology* **59**, 1209 (2014).

677. Qian, K. P., Natov, S. N., Pereira, B. J. & Lau, J. Y. Hepatitis C virus mixed genotype infection in patients on haemodialysis. *J Viral Hepat* **7**, 153–160 (2000).

678. Wu, J. C., Huang, I. A., Huang, Y. H., Chen, J. Y. & Sheen, I. J. Mixed genotypes infection with hepatitis D virus. *J Med Virol* **57**, 64–67 (1999).

679. Choisy, M. & de Roode, J. C. Mixed infections and the evolution of virulence: effects of resource competition, parasite plasticity, and impaired host immunity. *Am Nat* **175**, E105-118 (2010).

680. De Roode, J. C., Read, A. F., Chan, B. H. K. & Mackinnon, M. J. Rodent malaria parasites suffer from the presence of conspecific clones in three-clone *Plasmodium chabaudi* infections. *Parasitology* **127**, 411–418 (2003).

681. Balmer, O., Stearns, S. C., Schötzau, A. & Brun, R. Intraspecific competition between co-infecting parasite strains enhances host survival in African trypanosomes. *Ecology* **90**, 3367–3378 (2009).

682. Duval, A., Opatowski, L. & Brisse, S. Defining genomic epidemiology thresholds for common-source bacterial outbreaks: a modelling study. *The Lancet Microbe* **4**, e349–e357 (2023).

683. Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L. & Goering, R. V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection* **24**, 350–354 (2018).

684. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115 (2012).

685. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).

686. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43-46 (2007).

687. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).

688. Life Technologies. DNA Fragment Analysis by Capillary Electrophoresis. (2014).

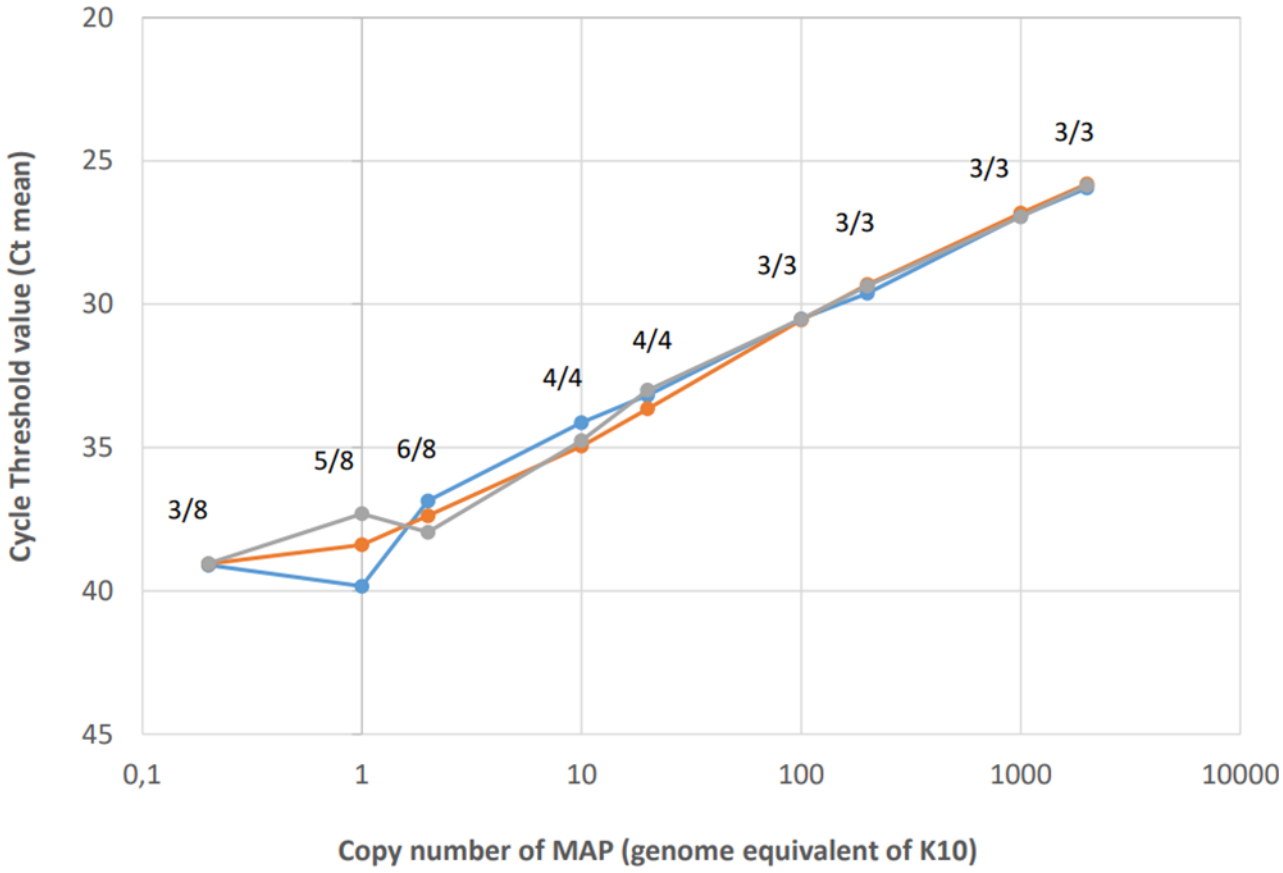
689. Fazekas, A., Steeves, R. & Newmaster, S. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* **48**, 277–285 (2010).

690. Kieleczawa, J. Fundamentals of Sequencing of Difficult Templates—An Overview. *J Biomol Tech* **17**, 207–217 (2006).

691. Sohal, J. S. *et al.* Genetic structure of *Mycobacterium avium* subsp. *paratuberculosis* population in cattle herds in Quebec as revealed by using a combination of multilocus genomic analyses. *J Clin Microbiol* **52**, 2764–2775 (2014).

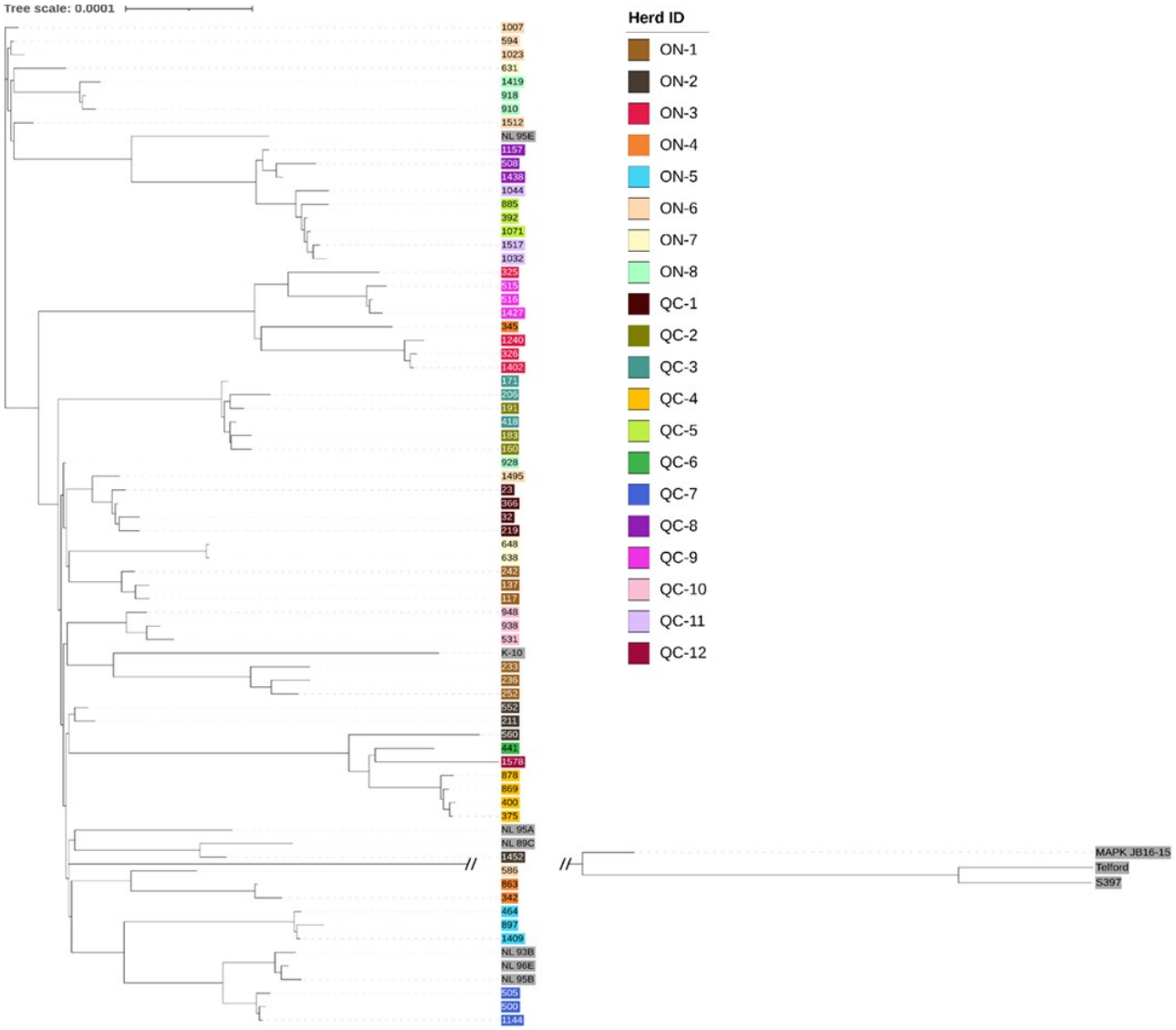
APPENDICES

Appendix I: Appendix Figures

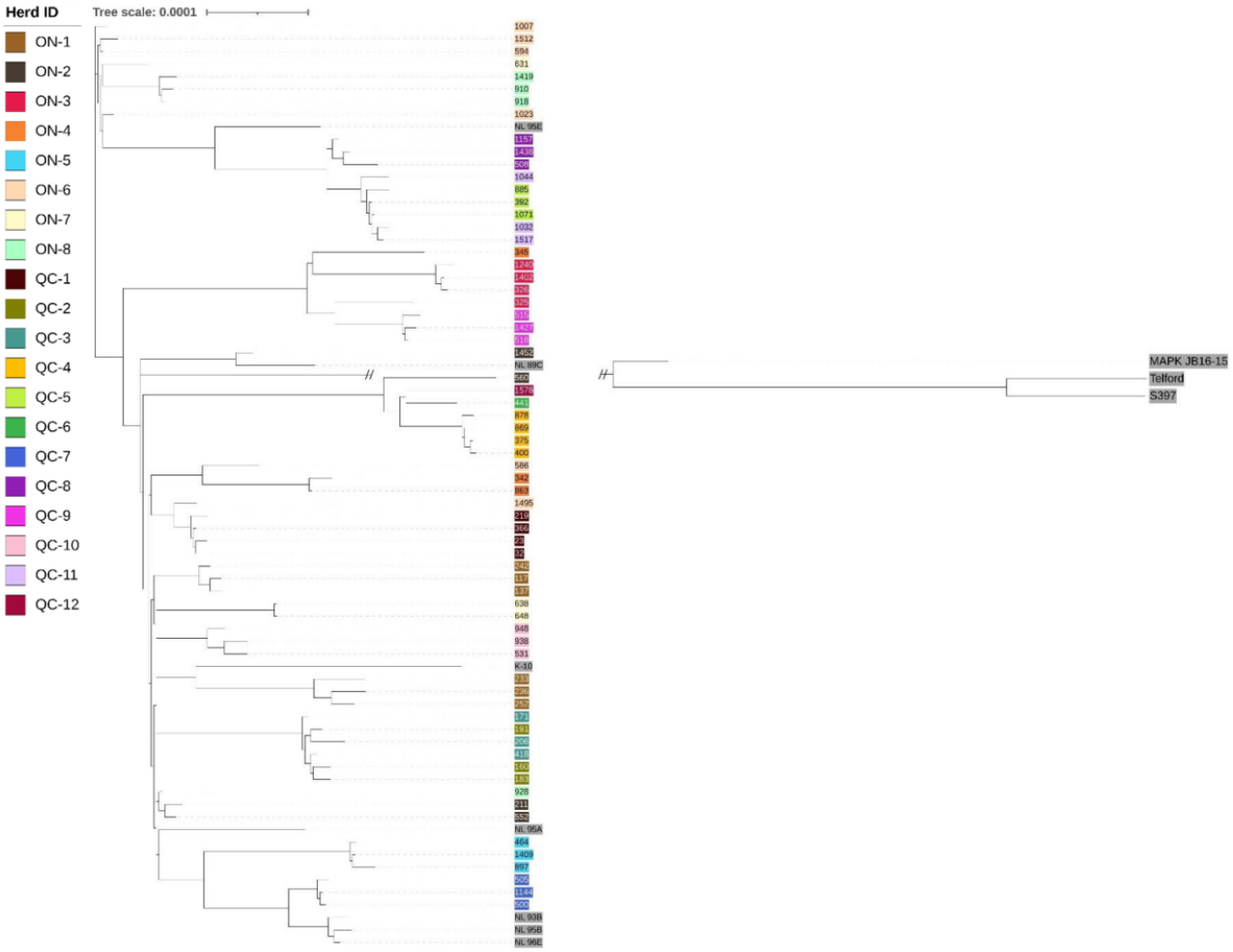


Appendix Figure A1: Detection of the K-10 strain of MAP by qPCR. Serial dilutions of DNA corresponding to known quantities of genome equivalents (Ge) were quantified by qPCR. Three independent serial dilution curves were made. Each dilution point was quantified eight times for the dilutions of 0.2, 1 and 2 Ge, four times for 10 and 20 Ge, and three times for the dilutions of 100, 200, 1000 and 2000 Ge.

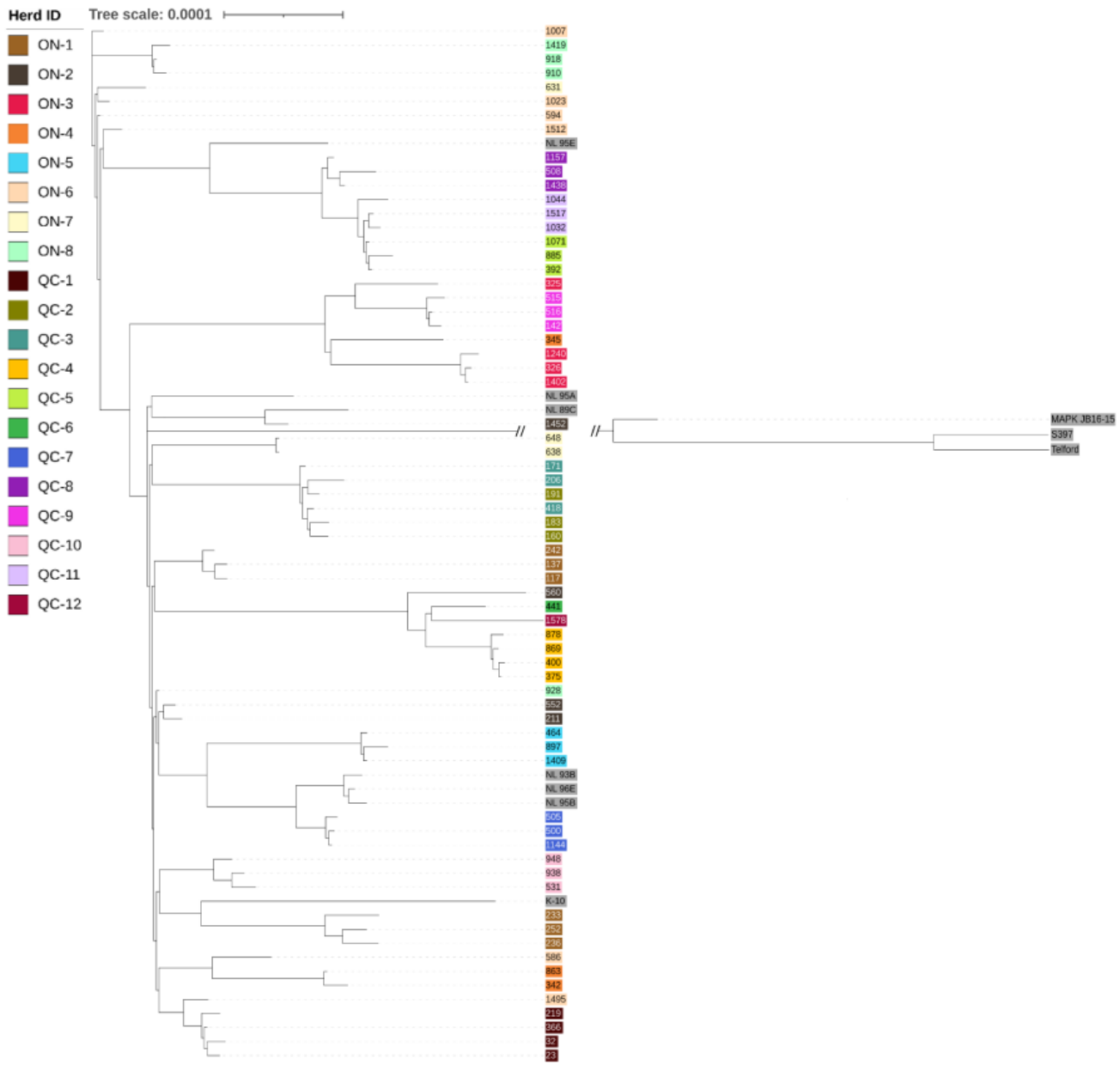
A)


Tree scale: 0.0001
1007
594
1023
631
1419
918
910
1512
NL 95E
1157
508
1438
1044
885
392
1071
1517
1032
325
315
516
1427
345
1240
326
1402
171
206
191
418
183
160
928
1495
23
366
32
219
648
638
242
137
117
948
938
531
K-10
283
236
252
552
211
560
441
1578
878
869
400
375
NL 95A
NL 89C
1452
586
863
342
464
897
1409
NL 93B
NL 96E
NL 95B
505
500
1144
MAPK JB16-15
Telford
S397
Herd ID
ON-1
ON-2
ON-3
ON-4
ON-5
ON-6
ON-7
ON-8
QC-1
QC-2
QC-3
QC-4
QC-5
QC-6
QC-7
QC-8
QC-9
QC-10
QC-11
QC-12

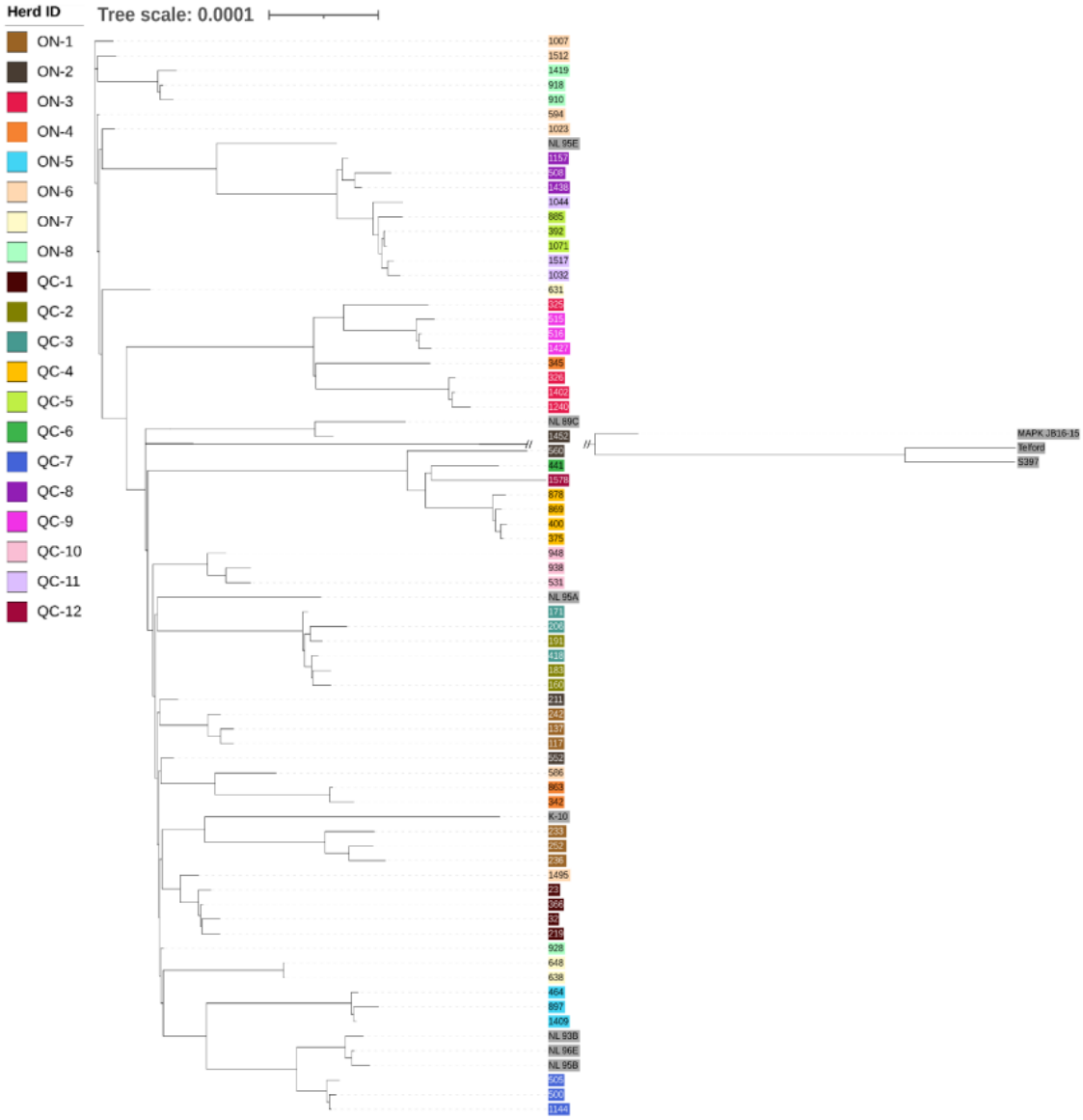
B)


Herd ID
Tree scale: 0.0001
ON-1
ON-2
ON-3
ON-4
ON-5
ON-6
ON-7
ON-8
QC-1
QC-2
QC-3
QC-4
QC-5
QC-6
QC-7
QC-8
QC-9
QC-10
QC-11
QC-12
MAPK JB16-15
Telford
S397

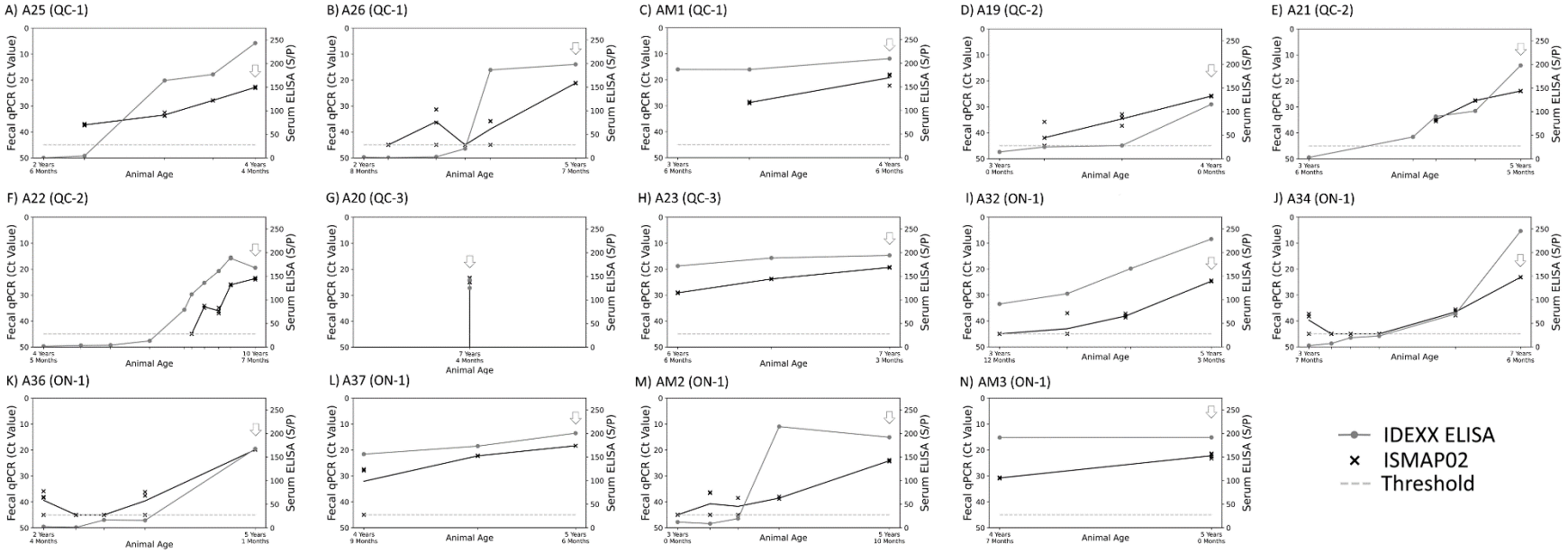
C)


Herd ID
Tree scale: 0.0001
ON-1
ON-2
ON-3
ON-4
ON-5
ON-6
ON-7
ON-8
QC-1
QC-2
QC-3
QC-4
QC-5
QC-6
QC-7
QC-8
QC-9
QC-10
QC-11
QC-12
1007
1419
918
910
631
1023
594
1512
NL 95E
1157
506
1438
1044
1517
1032
1071
885
392
325
515
516
142
345
1240
326
1402
NL 95A
NL 89C
1452
648
638
171
206
191
418
183
160
242
137
117
560
441
1578
878
869
400
375
928
552
211
464
897
1409
NL 93B
NL 96E
NL 95B
505
500
1144
948
938
531
K-10
233
252
236
586
863
342
1495
219
366
32
23
MAPK JB16-15
S397
Telford

D)

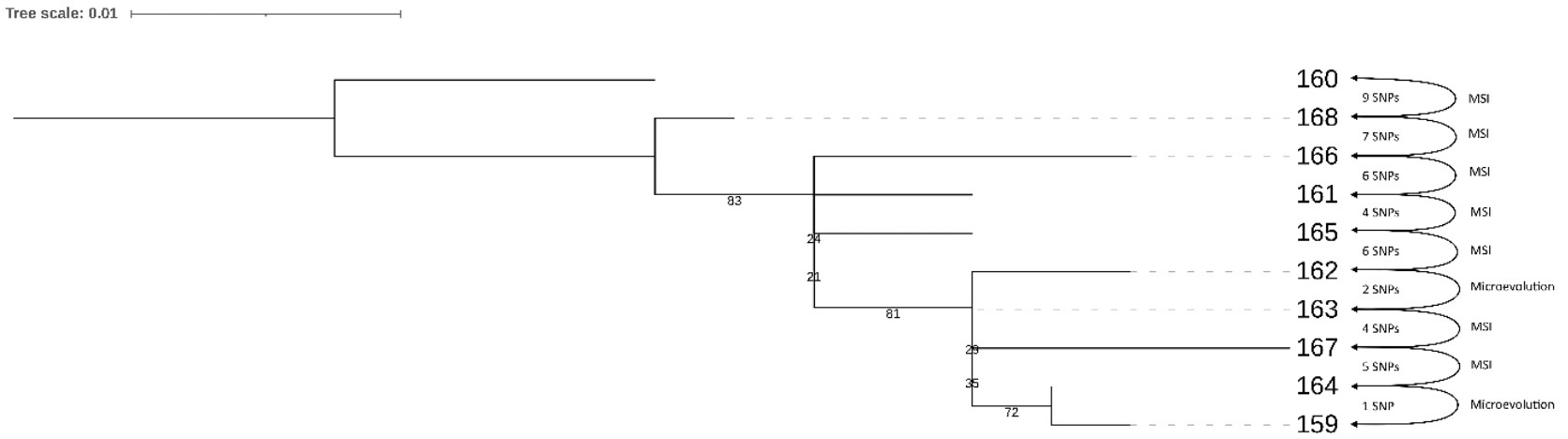


Appendix Figure A2: Core SNP phylogenies of MAP isolates (n=67) from ON and QC herds along with select reference sequences from NCBI (n=10). Each phylogeny was constructed based on 7 four separate core SNP analyses, each done using a different reference strain. A) used the type I reference strain Telford (NZ_CP033688.1), B) used the type II reference strain K-10 (NC_002944.2), C) used the type III reference strain S397 (NZ_CP053749.1), and D) used the type B reference strain MAPK_JB16/15 (NZ_CP033911.1). Each of the 67 field isolates are coloured by herd according to the listed key, while NCBI reference sequences are all coloured in grey.

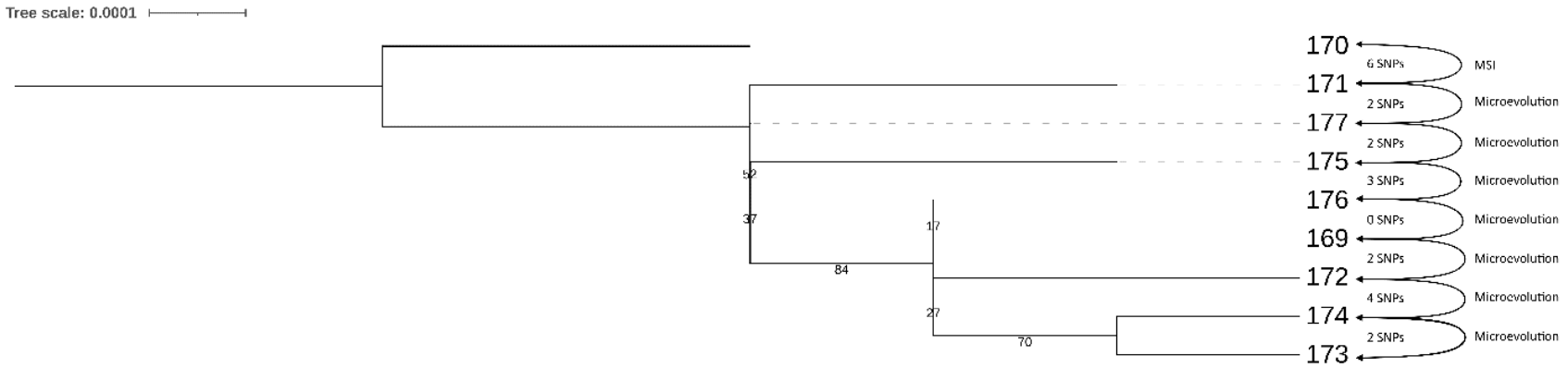


Appendix Figure A3: **Graphical plots depicting the relationship of animal age to both shedding of MAP in feces (qPCR C_t Value) and blood ELISA (% S/P) in all 14 animals examined in this study.** All animals were classified as high-shedding animals ($C_t < 27$) and the arrows on each graph represent the time point that fecal samples were collected for MAP culture in each animal. The grey line with "O" points represents the % S/P values of the serum ELISA. The black line with "X" points represents the C_t values obtained from fecal qPCR analysis. The dashed lines along the bottom of each graph represent the ELISA threshold (45% S/P) for a cow to be suspected of being MAP-infected positive (light gray), and the ELISA threshold (55% S/P) for a cow to be considered JD positive, as recommended by the manufacturer. Data collected for these graphs was provided by Dr. Bissonnette and AAFC.

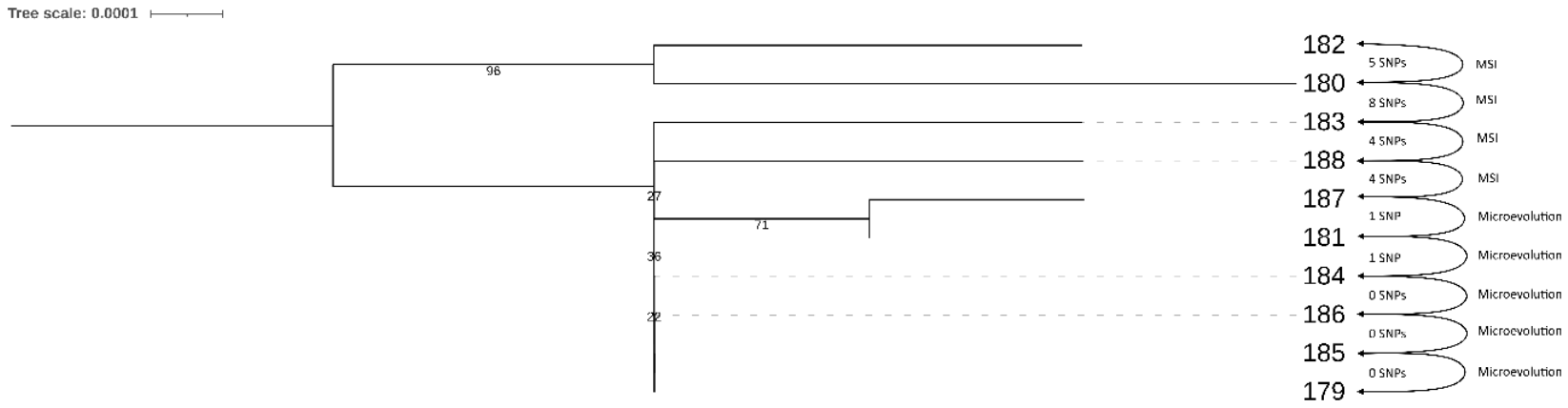
A A19 Core SNP Phylogeny (>2 SNPs required for MSI classification)



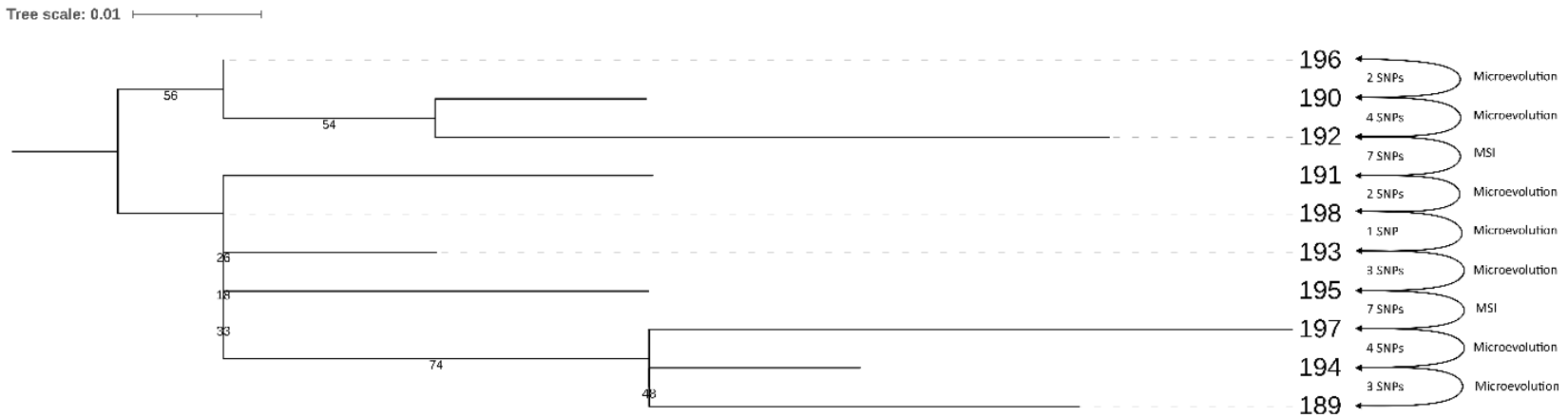
B A20 Core SNP Phylogeny (>4 SNPs required for MSI classification)



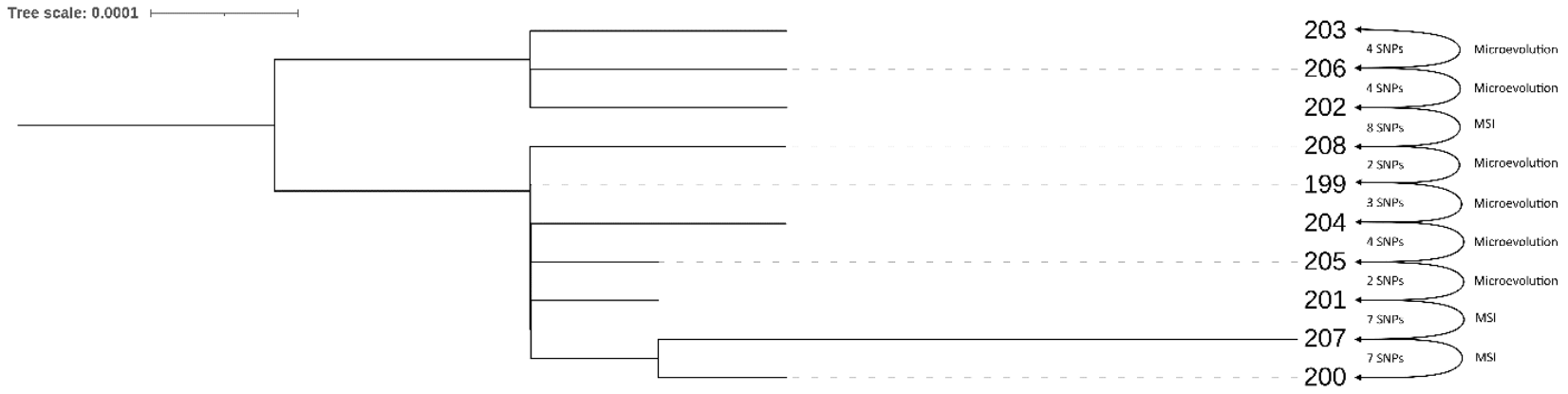
C A21 Core SNP Phylogeny (>3 SNPs required for MSI classification)



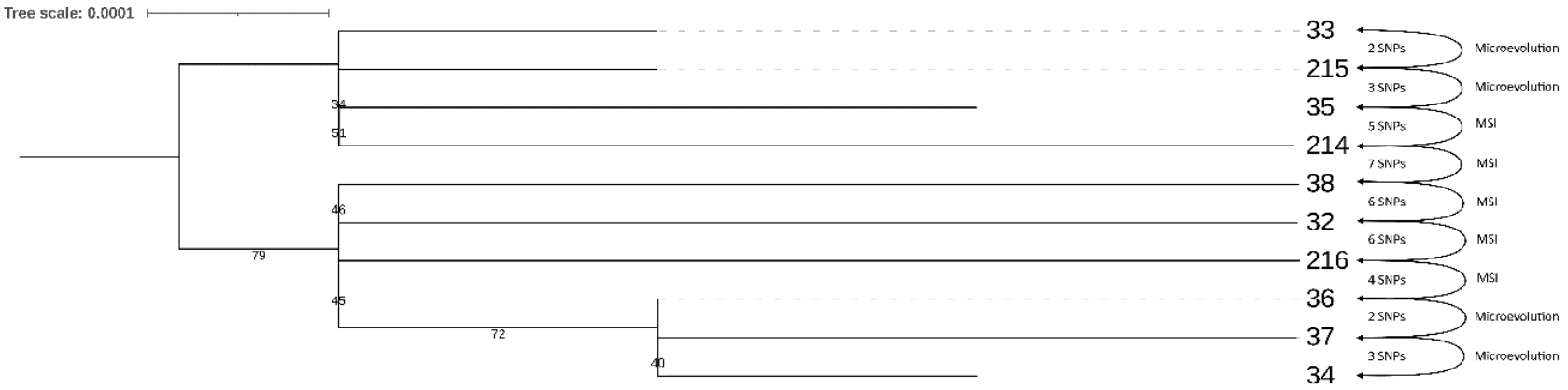
D A22 Core SNP Phylogeny (>6 SNPs required for MSI classification)



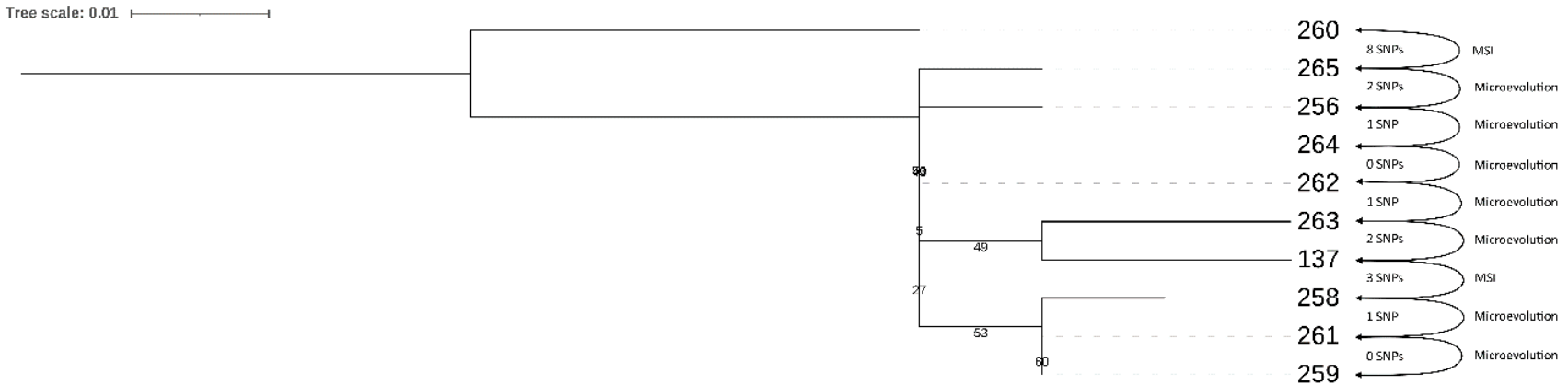
E A23 Core SNP Phylogeny (>4 SNPs required for MSI classification)



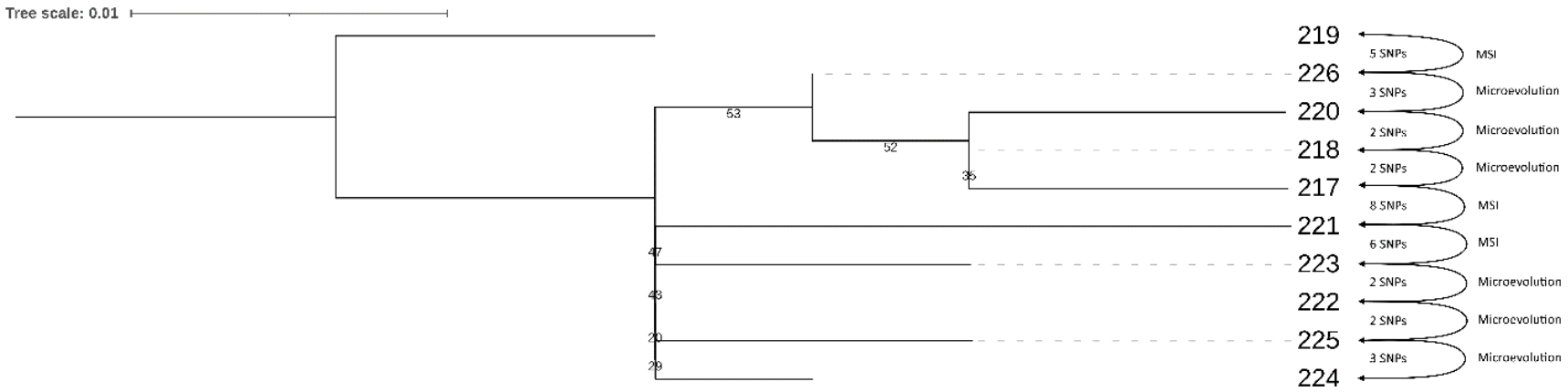
F A26 Core SNP Phylogeny (>3 SNPs required for MSI classification)



G A37 Core SNP Phylogeny (>3 SNPs required for MSI classification)



H AM1 Core SNP Phylogeny (>3 SNPs required for MSI classification)



I AM1 Core SNP Phylogeny (>3 SNPs required for MSI classification)

Figure A4: Phylogenetic trees constructed on a per-animal basis. The SNPs for all isolates from a single animal (n=10 for all except A20, where n=9) were used for tree construction by IQ-TREE and visualization using iTOL V6. A) Phylogeny for A19, B) Phylogeny for A20, C) Phylogeny for A21, D) Phylogeny for A22, E) Phylogeny for A23, F) Phylogeny for A26, G) Phylogeny for A37, H) Phylogeny for AM1, I) Phylogeny for AM2.

Appendix II: Appendix Tables

Appendix Table A1: Repeat Sizes of SSR1 and SSR2 for 1621 isolates retrieved from our personal collection and publicly available datasets as examined in Chapter V

Data Source[a]	Sequence Identifier[b]	Sample Name	Country of Origin	Continent of Origin	SSR1 Repeat Size	SSR1 Reading Frame[c]	SSR2 Repeat Size	SSR2 Reading Frame[d]
NCBI Nucleotide Database	GCA_000504725.1	08-8281	USA	North America	10	RF-1	9	RF-3
NCBI Nucleotide Database	GCA_000504785.1	10-4404	USA	North America	6	RF-3	9	RF-3
NCBI Nucleotide Database	GCA_000504845.1	10-5864	USA	North America	10	RF-1	10	RF-1
NCBI Nucleotide Database	GCA_000504865.1	10-5975	USA	North America	7	RF-1	8	RF-2
NCBI Nucleotide Database	GCA_000504885.1	10-8425	USA	North America	8	RF-2	UNK	UNK
NCBI Nucleotide Database	GCA_000504995.1	11-1786	USA	North America	UNK	UNK	9	RF-3
NCBI Nucleotide Database	GCA_001904575.1	2015WD-1	China	Asia	7	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_001904595.1	2015WD-2	China	Asia	7	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_016749935.1	42-13-1	Japan	Asia	9	RF-3	10	RF-1
NCBI Nucleotide Database	GCA_000240425.2	4B	USA	North America	7	RF-1	12	RF-3
NCBI Nucleotide Database	GCA_002180005.1	A3	Portugal	Europe	20	RF-2	11	RF-2
NCBI Nucleotide Database	GCA_000240525.2	ATCC 19698	USA	North America	7	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_002174295.1	C4A4	Portugal	Europe	8	RF-2	12	RF-3
NCBI Nucleotide Database	GCA_013357385.1	DSM 44135	Germany	Europe	7	RF-1	9	RF-3
NCBI Nucleotide Database	GCA_000240445.2	DT 3	USA	North America	11	RF-2	UNK	UNK
NCBI Nucleotide Database	GCA_000835225.1	E1	Egypt	Africa	17	RF-2	10	RF-1
NCBI Nucleotide Database	GCA_000835265.1	E93	Egypt	Africa	19	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_000240425.2	ENV 210	USA	North America	7	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_002211525.1	JII-1961	Germany	Europe	10	RF-1	10	RF-1
NCBI Nucleotide Database	GCA_002232005.2	JIII-386	Germany	Europe	7	RF-1	13	RF-1
NCBI Nucleotide Database	GCA_000240345.2	JQ5	Saudi Arabia	Asia	22	RF-1	15	RF-3
NCBI Nucleotide Database	GCA_000240365.2	JQ6	Saudi Arabia	Asia	16	RF-1	15	RF-3

NCBI Nucleotide Database	GCA_000240385.2	JTC 1281	USA	North America	7	RF-1	8	RF-2
NCBI Nucleotide Database	GCA_000240405.2	JTC 1285	USA	North America	12	RF-3	11	RF-2
NCBI Nucleotide Database	GCA_000007865.1	K-10	USA	North America	19	RF-1	10	RF-1
NCBI Nucleotide Database	GCA_001653355.1	MAP/TANUVAS/TN/India/2008	India	Asia	19	RF-1	10	RF-1
NCBI Nucleotide Database	GCA_000390085.1	MAP4	-	-	7	RF-1	12	RF-3
NCBI Nucleotide Database	GCA_003815795.1	MAPK_CN4/13	South Korea	Asia	7	RF-1	9	RF-3
NCBI Nucleotide Database	GCA_003713025.1	MAPK_CN7/15	South Korea	Asia	7	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_003713045.1	MAPK_CN9/15	South Korea	Asia	13	RF-1	9	RF-3
NCBI Nucleotide Database	GCA_003815815.1	MAPK_JB16/15	South Korea	Asia	7	RF-1	12	RF-3
NCBI Nucleotide Database	GCA_003816035.1	MAPK_JJ1/13	South Korea	Asia	7	RF-1	11	RF-2
NCBI Nucleotide Database	GCA_000218115.2	Pt154	Australia	Oceania	UNK	UNK	10	RF-1
NCBI Nucleotide Database	GCA_000219085.3	S397	USA	North America	5	RF-2	13	RF-1
NCBI Nucleotide Database	GCA_000330785.1	S5	India	Asia	7	RF-1	9	RF-3
NCBI Nucleotide Database/SRA Database	GCA_000218035.2, SRR136596	Pt139	Australia	Oceania	20	RF-2	10	RF-1
NCBI Nucleotide Database/SRA Database	GCA_000218055.2, SRR136755	Pt144	Australia	Oceania	UNK	UNK	10	RF-1
NCBI Nucleotide Database/SRA Database	GCA_000218075.2, SRR197999	Pt145	Australia	Oceania	UNK	UNK	10	RF-1
NCBI Nucleotide Database/SRA Database	GCA_000218095.2, SRR198001	Pt146	Australia	Oceania	19	RF-1	11	RF-2
NCBI Nucleotide Database/SRA Database	GCA_000218135.2, SRR201792	Pt155	Australia	Oceania	UNK	UNK	10	RF-1
NCBI Nucleotide Database/SRA Database	GCA_000218155.2, SRR198003	Pt164	Australia	Oceania	UNK	UNK	10	RF-1
QC/ON Isolate WGS	SRR23179856	AAFC_MAP_23	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326312	AAFC_MAP_#24	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326311	AAFC_MAP_#25	Canada	North America	16	RF-1	9	RF-3
QC/ON Isolate WGS	SRR24326254	AAFC_MAP_#26	Canada	North America	15	RF-3	9	RF-3
QC/ON Isolate WGS	SRR24326226	AAFC_MAP_#27	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326214	AAFC_MAP_#28	Canada	North America	13	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326204	AAFC_MAP_#29	Canada	North America	16	RF-1	12	RF-3
QC/ON Isolate WGS	SRR24326193	AAFC_MAP_#30	Canada	North America	11	RF-2	10	RF-1

QC/ON Isolate WGS	SRR23179855	AAFC_MAP_32	Canada	North America	11	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326239	AAFC_MAP_#33	Canada	North America	13	RF-1	9	RF-3
QC/ON Isolate WGS	SRR24326310	AAFC_MAP_#34	Canada	North America	13	RF-1	12	RF-3
QC/ON Isolate WGS	SRR24326298	AAFC_MAP_#35	Canada	North America	11	RF-2	11	RF-2
QC/ON Isolate WGS	SRR24326267	AAFC_MAP_#36	Canada	North America	11	RF-2	11	RF-2
QC/ON Isolate WGS	SRR24326261	AAFC_MAP_#37	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326260	AAFC_MAP_#38	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326228	AAFC_MAP_#113	Canada	North America	18	RF-3	14	RF-2
QC/ON Isolate WGS	SRR24326248	AAFC_MAP_#114	Canada	North America	17	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326227	AAFC_MAP_#115	Canada	North America	14	RF-2	UNK	UNK
QC/ON Isolate WGS	SRR24326225	AAFC_MAP_#116	Canada	North America	13	RF-1	14	RF-2
QC/ON Isolate WGS	SRR23179835	AAFC_MAP_117	Canada	North America	15	RF-3	13	RF-1
QC/ON Isolate WGS	SRR24326224	AAFC_MAP_#118	Canada	North America	17	RF-2	15	RF-3
QC/ON Isolate WGS	SRR24326222	AAFC_MAP_#119	Canada	North America	18	RF-3	16	RF-1
QC/ON Isolate WGS	SRR24326221	AAFC_MAP_#120	Canada	North America	18	RF-3	18	RF-3
QC/ON Isolate WGS	SRR24326220	AAFC_MAP_#121	Canada	North America	14	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326223	AAFC_MAP_#122	Canada	North America	15	RF-3	12	RF-3
QC/ON Isolate WGS	SRR24326219	AAFC_MAP_#123	Canada	North America	14	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326218	AAFC_MAP_#124	Canada	North America	19	RF-1	14	RF-2
QC/ON Isolate WGS	SRR24326217	AAFC_MAP_#125	Canada	North America	17	RF-2	15	RF-3
QC/ON Isolate WGS	SRR24326209	AAFC_MAP_#126	Canada	North America	13	RF-1	13	RF-1
QC/ON Isolate WGS	SRR24326208	AAFC_MAP_#127	Canada	North America	19	RF-1	11	RF-2
QC/ON Isolate WGS	SRR24326207	AAFC_MAP_#128	Canada	North America	19	RF-1	13	RF-1
QC/ON Isolate WGS	SRR24326206	AAFC_MAP_#129	Canada	North America	17	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326205	AAFC_MAP_#130	Canada	North America	14	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326203	AAFC_MAP_#131	Canada	North America	14	RF-2	14	RF-2

QC/ON Isolate WGS	SRR24326202	AAFC_MAP_#132	Canada	North America	22	RF-1	11	RF-2
QC/ON Isolate WGS	SRR24326201	AAFC_MAP_#133	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326199	AAFC_MAP_#134	Canada	North America	20	RF-2	15	RF-3
QC/ON Isolate WGS	SRR24326198	AAFC_MAP_#135	Canada	North America	17	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326188	AAFC_MAP_#136	Canada	North America	UNK	UNK	13	RF-1
QC/ON Isolate WGS	SRR23179791	AAFC_MAP_137	Canada	North America	20	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326270	AAFC_MAP_#159	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR23179790	AAFC_MAP_160	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326268	AAFC_MAP_#161	Canada	North America	15	RF-3	9	RF-3
QC/ON Isolate WGS	SRR24326246	AAFC_MAP_#162	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326245	AAFC_MAP_#163	Canada	North America	UNK	UNK	10	RF-1
QC/ON Isolate WGS	SRR24326269	AAFC_MAP_#164	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326244	AAFC_MAP_#165	Canada	North America	12	RF-3	13	RF-1
QC/ON Isolate WGS	SRR24326243	AAFC_MAP_#166	Canada	North America	14	RF-2	9	RF-3
QC/ON Isolate WGS	SRR24326242	AAFC_MAP_#167	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326241	AAFC_MAP_#168	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326238	AAFC_MAP_#169	Canada	North America	17	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326237	AAFC_MAP_#170	Canada	North America	UNK	UNK	14	RF-2
QC/ON Isolate WGS	SRR23179854	AAFC_MAP_171	Canada	North America	17	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326240	AAFC_MAP_#172	Canada	North America	15	RF-3	13	RF-1
QC/ON Isolate WGS	SRR24326236	AAFC_MAP_#173	Canada	North America	15	RF-3	13	RF-1
QC/ON Isolate WGS	SRR24326235	AAFC_MAP_#174	Canada	North America	14	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326234	AAFC_MAP_#175	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326233	AAFC_MAP_#176	Canada	North America	21	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326232	AAFC_MAP_#177	Canada	North America	12	RF-3	12	RF-3
QC/ON Isolate WGS	SRR24326231	AAFC_MAP_#179	Canada	North America	16	RF-1	10	RF-1

QC/ON Isolate WGS	SRR24326230	AAFC_MAP_#180	Canada	North America	20	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326309	AAFC_MAP_#181	Canada	North America	23	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326307	AAFC_MAP_#182	Canada	North America	15	RF-3	14	RF-2
QC/ON Isolate WGS	SRR23179853	AAFC_MAP_183	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326306	AAFC_MAP_#184	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326305	AAFC_MAP_#185	Canada	North America	18	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326308	AAFC_MAP_#186	Canada	North America	17	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326304	AAFC_MAP_#187	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326303	AAFC_MAP_#188	Canada	North America	19	RF-1	12	RF-3
QC/ON Isolate WGS	SRR24326302	AAFC_MAP_#189	Canada	North America	14	RF-2	14	RF-2
QC/ON Isolate WGS	SRR24326301	AAFC_MAP_#190	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR23179851	AAFC_MAP_191	Canada	North America	14	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326299	AAFC_MAP_#192	Canada	North America	17	RF-2	11	RF-2
QC/ON Isolate WGS	SRR24326297	AAFC_MAP_#193	Canada	North America	14	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326300	AAFC_MAP_#194	Canada	North America	16	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326296	AAFC_MAP_#195	Canada	North America	14	RF-2	12	RF-3
QC/ON Isolate WGS	SRR24326295	AAFC_MAP_#196	Canada	North America	16	RF-1	11	RF-2
QC/ON Isolate WGS	SRR24326294	AAFC_MAP_#197	Canada	North America	19	RF-1	13	RF-1
QC/ON Isolate WGS	SRR24326293	AAFC_MAP_#198	Canada	North America	13	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326291	AAFC_MAP_#199	Canada	North America	15	RF-3	15	RF-3
QC/ON Isolate WGS	SRR24326290	AAFC_MAP_#200	Canada	North America	19	RF-1	15	RF-3
QC/ON Isolate WGS	SRR24326289	AAFC_MAP_#201	Canada	North America	12	RF-3	18	RF-3
QC/ON Isolate WGS	SRR24326292	AAFC_MAP_#202	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326266	AAFC_MAP_#203	Canada	North America	14	RF-2	13	RF-1
QC/ON Isolate WGS	SRR24326265	AAFC_MAP_#204	Canada	North America	15	RF-3	16	RF-1
QC/ON Isolate WGS	SRR24326264	AAFC_MAP_#205	Canada	North America	13	RF-1	15	RF-3

QC/ON Isolate WGS	SRR23179850	AAFC_MAP_206	Canada	North America	15	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326263	AAFC_MAP_#207	Canada	North America	14	RF-2	14	RF-2
QC/ON Isolate WGS	SRR24326262	AAFC_MAP_#208	Canada	North America	12	RF-3	12	RF-3
QC/ON Isolate WGS	SRR23179849	AAFC_MAP_211	Canada	North America	18	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326283	AAFC_MAP_#212	Canada	North America	19	RF-1	9	RF-3
QC/ON Isolate WGS	SRR24326272	AAFC_MAP_#213	Canada	North America	20	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326259	AAFC_MAP_#214	Canada	North America	10	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326258	AAFC_MAP_#215	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326257	AAFC_MAP_#216	Canada	North America	16	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326256	AAFC_MAP_#217	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326255	AAFC_MAP_#218	Canada	North America	14	RF-2	10	RF-1
QC/ON Isolate WGS	SRR23179843	AAFC_MAP_219	Canada	North America	11	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326253	AAFC_MAP_#220	Canada	North America	11	RF-2	11	RF-2
QC/ON Isolate WGS	SRR24326252	AAFC_MAP_#221	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326251	AAFC_MAP_#222	Canada	North America	12	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326250	AAFC_MAP_#223	Canada	North America	12	RF-3	12	RF-3
QC/ON Isolate WGS	SRR24326249	AAFC_MAP_#224	Canada	North America	11	RF-2	11	RF-2
QC/ON Isolate WGS	SRR24326247	AAFC_MAP_#225	Canada	North America	14	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326229	AAFC_MAP_#226	Canada	North America	15	RF-3	12	RF-3
QC/ON Isolate WGS	SRR24326216	AAFC_MAP_#228	Canada	North America	18	RF-3	11	RF-2
QC/ON Isolate WGS	SRR24326213	AAFC_MAP_#229	Canada	North America	15	RF-3	14	RF-2
QC/ON Isolate WGS	SRR24326212	AAFC_MAP_#230	Canada	North America	20	RF-2	14	RF-2
QC/ON Isolate WGS	SRR24326215	AAFC_MAP_#231	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326211	AAFC_MAP_#232	Canada	North America	20	RF-2	12	RF-3
QC/ON Isolate WGS	SRR23179800	AAFC_MAP_233	Canada	North America	19	RF-1	13	RF-1
QC/ON Isolate WGS	SRR24326210	AAFC_MAP_#234	Canada	North America	18	RF-3	13	RF-1

QC/ON Isolate WGS	SRR23179829	AAFC_MAP_236	Canada	North America	14	RF-2	13	RF-1
QC/ON Isolate WGS	SRR24326200	AAFC_MAP_#237	Canada	North America	14	RF-2	13	RF-1
QC/ON Isolate WGS	SRR24326196	AAFC_MAP_#238	Canada	North America	UNK	UNK	14	RF-2
QC/ON Isolate WGS	SRR24326195	AAFC_MAP_#239	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326194	AAFC_MAP_#240	Canada	North America	20	RF-2	11	RF-2
QC/ON Isolate WGS	SRR24326197	AAFC_MAP_#241	Canada	North America	16	RF-1	11	RF-2
QC/ON Isolate WGS	SRR23179819	AAFC_MAP_242	Canada	North America	22	RF-1	13	RF-1
QC/ON Isolate WGS	SRR24326192	AAFC_MAP_#243	Canada	North America	17	RF-2	13	RF-1
QC/ON Isolate WGS	SRR24326191	AAFC_MAP_#244	Canada	North America	22	RF-1	14	RF-2
QC/ON Isolate WGS	SRR24326190	AAFC_MAP_#245	Canada	North America	22	RF-1	13	RF-1
QC/ON Isolate WGS	SRR24326288	AAFC_MAP_#247	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326287	AAFC_MAP_#248	Canada	North America	UNK	UNK	12	RF-3
QC/ON Isolate WGS	SRR24326189	AAFC_MAP_#249	Canada	North America	24	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326286	AAFC_MAP_#250	Canada	North America	UNK	UNK	11	RF-2
QC/ON Isolate WGS	SRR24326285	AAFC_MAP_#251	Canada	North America	UNK	UNK	13	RF-1
QC/ON Isolate WGS	SRR23179792	AAFC_MAP_252	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326284	AAFC_MAP_#253	Canada	North America	UNK	UNK	12	RF-3
QC/ON Isolate WGS	SRR24326282	AAFC_MAP_#254	Canada	North America	20	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326281	AAFC_MAP_#255	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326280	AAFC_MAP_#256	Canada	North America	UNK	UNK	11	RF-2
QC/ON Isolate WGS	SRR24326279	AAFC_MAP_#258	Canada	North America	UNK	UNK	10	RF-1
QC/ON Isolate WGS	SRR24326278	AAFC_MAP_#259	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR24326276	AAFC_MAP_#260	Canada	North America	UNK	UNK	13	RF-1
QC/ON Isolate WGS	SRR24326275	AAFC_MAP_#261	Canada	North America	17	RF-2	10	RF-1
QC/ON Isolate WGS	SRR24326274	AAFC_MAP_#262	Canada	North America	13	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326277	AAFC_MAP_#263	Canada	North America	19	RF-1	10	RF-1

QC/ON Isolate WGS	SRR24326273	AAFC_MAP_#264	Canada	North America	16	RF-1	10	RF-1
QC/ON Isolate WGS	SRR24326271	AAFC_MAP_#265	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179852	AAFC_MAP_325	Canada	North America	7	RF-1	11	RF-2
QC/ON Isolate WGS	SRR23179848	AAFC_MAP_326	Canada	North America	7	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179845	AAFC_MAP_342	Canada	North America	9	RF-3	UNK	UNK
QC/ON Isolate WGS	SRR23179842	AAFC_MAP_345	Canada	North America	7	RF-1	12	RF-3
QC/ON Isolate WGS	SRR23179841	AAFC_MAP_366	Canada	North America	12	RF-3	10	RF-1
QC/ON Isolate WGS	SRR23179840	AAFC_MAP_375	Canada	North America	18	RF-3	10	RF-1
QC/ON Isolate WGS	SRR23179837	AAFC_MAP_392	Canada	North America	7	RF-1	9	RF-3
QC/ON Isolate WGS	SRR23179838	AAFC_MAP_400	Canada	North America	20	RF-2	11	RF-2
QC/ON Isolate WGS	SRR23179811	AAFC_MAP_418	Canada	North America	20	RF-2	10	RF-1
QC/ON Isolate WGS	SRR23179810	AAFC_MAP_441	Canada	North America	14	RF-2	9	RF-3
QC/ON Isolate WGS	SRR23179807	AAFC_MAP_464	Canada	North America	15	RF-3	12	RF-3
QC/ON Isolate WGS	SRR23179799	AAFC_MAP_500	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179798	AAFC_MAP_505	Canada	North America	19	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179795	AAFC_MAP_508	Canada	North America	7	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179836	AAFC_MAP_515	Canada	North America	9	RF-3	11	RF-2
QC/ON Isolate WGS	SRR23179833	AAFC_MAP_516	Canada	North America	9	RF-3	13	RF-1
QC/ON Isolate WGS	SRR23179831	AAFC_MAP_531	Canada	North America	21	RF-3	10	RF-1
QC/ON Isolate WGS	SRR23179828	AAFC_MAP_552	Canada	North America	13	RF-1	9	RF-3
QC/ON Isolate WGS	SRR23179827	AAFC_MAP_560	Canada	North America	11	RF-2	9	RF-3
QC/ON Isolate WGS	SRR23179826	AAFC_MAP_586	Canada	North America	14	RF-2	12	RF-3
QC/ON Isolate WGS	SRR23179825	AAFC_MAP_594	Canada	North America	7	RF-1	13	RF-1
QC/ON Isolate WGS	SRR23179815	AAFC_MAP_631	Canada	North America	7	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179814	AAFC_MAP_638	Canada	North America	20	RF-2	10	RF-1
QC/ON Isolate WGS	SRR23179813	AAFC_MAP_648	Canada	North America	UNK	UNK	10	RF-1

QC/ON Isolate WGS	SRR23179846	AAFC_MAP_863	Canada	North America	9	RF-3	17	RF-2
QC/ON Isolate WGS	SRR23179844	AAFC_MAP_869	Canada	North America	13	RF-1	11	RF-2
QC/ON Isolate WGS	SRR23179839	AAFC_MAP_878	Canada	North America	15	RF-3	16	RF-1
QC/ON Isolate WGS	SRR23179812	AAFC_MAP_885	Canada	North America	7	RF-1	9	RF-3
QC/ON Isolate WGS	SRR23179809	AAFC_MAP_897	Canada	North America	21	RF-3	10	RF-1
QC/ON Isolate WGS	SRR23179806	AAFC_MAP_910	Canada	North America	7	RF-1	13	RF-1
QC/ON Isolate WGS	SRR23179805	AAFC_MAP_918	Canada	North America	7	RF-1	11	RF-2
QC/ON Isolate WGS	SRR23179803	AAFC_MAP_928	Canada	North America	11	RF-2	9	RF-3
QC/ON Isolate WGS	SRR23179802	AAFC_MAP_938	Canada	North America	18	RF-3	12	RF-3
QC/ON Isolate WGS	SRR23179801	AAFC_MAP_948	Canada	North America	11	RF-2	17	RF-2
QC/ON Isolate WGS	SRR23179823	AAFC_MAP_1007	Canada	North America	7	RF-1	12	RF-3
QC/ON Isolate WGS	SRR23179822	AAFC_MAP_1023	Canada	North America	7	RF-1	14	RF-2
QC/ON Isolate WGS	SRR23179820	AAFC_MAP_1032	Canada	North America	7	RF-1	9	RF-3
QC/ON Isolate WGS	SRR23179817	AAFC_MAP_1044	Canada	North America	7	RF-1	11	RF-2
QC/ON Isolate WGS	SRR23179834	AAFC_MAP_1071	Canada	North America	7	RF-1	9	RF-3
QC/ON Isolate WGS	SRR23179797	AAFC_MAP_1144	Canada	North America	18	RF-3	9	RF-3
QC/ON Isolate WGS	SRR23179796	AAFC_MAP_1157	Canada	North America	7	RF-1	13	RF-1
QC/ON Isolate WGS	SRR23179816	AAFC_MAP_1240	Canada	North America	7	RF-1	14	RF-2
QC/ON Isolate WGS	SRR23179847	AAFC_MAP_1402	Canada	North America	7	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179808	AAFC_MAP_1409	Canada	North America	11	RF-2	12	RF-3
QC/ON Isolate WGS	SRR23179804	AAFC_MAP_1419	Canada	North America	7	RF-1	11	RF-2
QC/ON Isolate WGS	SRR23179794	AAFC_MAP_1427	Canada	North America	10	RF-1	12	RF-3
QC/ON Isolate WGS	SRR23179830	AAFC_MAP_1438	Canada	North America	7	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179832	AAFC_MAP_1452	Canada	North America	7	RF-1	10	RF-1
QC/ON Isolate WGS	SRR23179821	AAFC_MAP_1495	Canada	North America	9	RF-3	11	RF-2
QC/ON Isolate WGS	SRR23179824	AAFC_MAP_1512	Canada	North America	7	RF-1	10	RF-1

QC/ON Isolate WGS	SRR23179818	AAFC_MAP_1517	Canada	North America	7	RF-1	9	RF-3
QC/ON Isolate WGS	SRR23179793	AAFC_MAP_1578	Canada	North America	10	RF-1	16	RF-1
SRA Database	SRR13214443	3324	Spain	Europe	UNK	UNK	11	RF-2
SRA Database	SRR13214442	3326	Spain	Europe	UNK	UNK	10	RF-1
SRA Database	SRR13214444	3410	Australia	Oceania	10	RF-1	10	RF-1
SRA Database	SRR13214445	3413	Australia	Oceania	8	RF-2	10	RF-1
SRA Database	SRR13214448	3443	Australia	Oceania	8	RF-2	12	RF-3
SRA Database	SRR19395037	005-1467	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19395036	005-1469	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR19394963	005-1473	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR19394952	005-8538	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394941	005-8626	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394930	005-8630	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR13214446	110b	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR18116652	110c	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR18116703	111a	Australia	Oceania	12	RF-3	12	RF-3
SRA Database	SRR18116702	111c	Australia	Oceania	12	RF-3	11	RF-2
SRA Database	SRR18116701	112a	Australia	Oceania	10	RF-1	10	RF-1
SRA Database	SRR18116699	112b	Australia	Oceania	12	RF-3	12	RF-3
SRA Database	SRR18116698	112c	Australia	Oceania	10	RF-1	12	RF-3
SRA Database	SRR18116686	115a	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR18116685	115b	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR18116684	115c	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR18116649	117b	Australia	Oceania	9	RF-3	13	RF-1
SRA Database	SRR18116648	117c	Australia	Oceania	9	RF-3	UNK	UNK
SRA Database	SRR18116663	118a	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR18116660	121b	Australia	Oceania	9	RF-3	12	RF-3
SRA Database	SRR18116659	121c	Australia	Oceania	9	RF-3	UNK	UNK
SRA Database	SRR18116657	123b	Australia	Oceania	UNK	UNK	UNK	UNK

SRA Database	SRR18116651	126a	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR18116650	126b	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR18116697	127a	Australia	Oceania	12	RF-3	12	RF-3
SRA Database	SRR18116696	127b	Australia	Oceania	11	RF-2	11	RF-2
SRA Database	SRR18116695	127c	Australia	Oceania	11	RF-2	11	RF-2
SRA Database	SRR18116683	128b	Australia	Oceania	11	RF-2	12	RF-3
SRA Database	SRR18116682	128c	Australia	Oceania	10	RF-1	11	RF-2
SRA Database	SRR18116681	129c	Australia	Oceania	9	RF-3	12	RF-3
SRA Database	SRR18116692	130a	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR18116691	130b	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR18116690	130c	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR14863475	13-2711	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR18116647	132a	Australia	Oceania	9	RF-3	12	RF-3
SRA Database	SRR18116646	132c	Australia	Oceania	10	RF-1	13	RF-1
SRA Database	SRR18116688	133b	Australia	Oceania	12	RF-3	11	RF-2
SRA Database	SRR18116687	133c	Australia	Oceania	12	RF-3	11	RF-2
SRA Database	SRR14863474	13-4284	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863366	13-4516	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863355	13-4573	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR12593860	1347_498_INMV2	Argentina	South America	7	RF-1	11	RF-2
SRA Database	SRR14863344	13-4959	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863333	13-4961	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863322	13-5083	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863311	13-5941	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863300	13-5950	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863289	13-5985	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR18116694	135a	Australia	Oceania	11	RF-2	UNK	UNK
SRA Database	SRR13214447	135b	Australia	Oceania	11	RF-2	UNK	UNK
SRA Database	SRR18116693	135c	Australia	Oceania	10	RF-1	10	RF-1

SRA Database	SRR14863473	13-6371	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863462	13-6806	Ireland	Europe	8	RF-2	12	RF-3
SRA Database	SRR14863451	13-6823	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863440	13-6914	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863429	13-8781	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863418	14-1252	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863407	14-2603	Ireland	Europe	UNK	UNK	13	RF-1
SRA Database	SRR14863396	14-2607	Ireland	Europe	15	RF-3	9	RF-3
SRA Database	SRR14863385	14-2622	Ireland	Europe	UNK	UNK	10	RF-1
SRA Database	SRR14863374	14-2629	Ireland	Europe	12	RF-3	11	RF-2
SRA Database	SRR14863365	14-3230	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863364	14-3720	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863363	14-3851	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863362	14-3861	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863361	14-3960	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863360	14-3966	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863359	14-4569	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863358	14-4946	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863357	14-5165	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863356	14-5172	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863354	14-5298	Ireland	Europe	8	RF-2	11	RF-2
SRA Database	SRR14863353	14-5300	Ireland	Europe	UNK	UNK	12	RF-3
SRA Database	SRR14863352	14-5600	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863351	14-5700	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863350	14-6194	Ireland	Europe	7	RF-1	8	RF-2
SRA Database	SRR14863349	14-6195	Ireland	Europe	7	RF-1	8	RF-2
SRA Database	SRR14863348	14-6254	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863347	14-6255	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863346	14-6342	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR14863345	14-6430	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863343	14-6538	Ireland	Europe	8	RF-2	11	RF-2
SRA Database	SRR14863342	14-6670	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863341	14-6677	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863340	14-7252	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863339	14-7486	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863338	14-7488	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863337	14-7489	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863336	14-7530	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863335	14-7534	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863334	14-7535	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863332	14-7626	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863331	14-7736	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863330	14-7739	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863329	14-8576	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863328	14-9252	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863327	14-9294	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR18116680	14a	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR18116679	14b	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR12593859	1543_481_INMV2	Argentina	South America	7	RF-1	11	RF-2
SRA Database	SRR14863326	16-1173	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863325	16-1297	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863324	16-1298	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863323	16-1734	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863321	16-1735	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863320	16-2363	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863319	16-2463	Ireland	Europe	8	RF-2	9	RF-3
SRA Database	SRR14863318	16-2468	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR14863317	16-2480	Ireland	Europe	8	RF-2	10	RF-1

SRA Database	SRR14863316	16-2495	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR14863315	16-2502	Ireland	Europe	8	RF-2	11	RF-2
SRA Database	SRR14863314	16-2504	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR14863313	16-2620	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863312	16-3075	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863310	16-3373	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863309	16-3374	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863308	16-406	Ireland	Europe	10	RF-1	11	RF-2
SRA Database	SRR14863307	16-4251	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863306	16-4796	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863305	16-4870	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863304	16-4934	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863303	16-4936	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863302	16-5154	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863301	16-5285	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR19394919	16-5548	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395007	16-5552	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863299	16-5561	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394996	16-5567	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394910	16-5577	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395035	16-5578	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863298	16-5757	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR14863297	16-5836	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863296	16-5947	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863295	16-6024	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863294	16-665	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863293	16-666	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863292	16-844	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863291	16-845	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR18116705	16a	Australia	Oceania	10	RF-1	13	RF-1
SRA Database	SRR18116704	16b	Australia	Oceania	10	RF-1	12	RF-3
SRA Database	SRR14863290	17-1627	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863288	17-1729	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863287	17-2127	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863286	17-2523	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863285	17-2658	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863284	17-2705	Ireland	Europe	10	RF-1	10	RF-1
SRA Database	SRR14863283	17-2827	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863282	17-2860	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863281	17-2963	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863280	17-3112	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863279	17-3451	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863472	17-3523	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863471	17-3831	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863470	17-4108	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863469	17-4774	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863468	17-4775	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863467	17-4776	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863466	17-5544	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863465	17-5546	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR19395024	17-5556	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394990	17-5559	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394978	17-5560	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394969	17-5561	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394968	17-5564	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394967	17-5566	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394966	17-5567	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394965	17-5569	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR19394964	17-5570	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394962	17-5571	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR19394961	17-5572	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR19394960	17-5583	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394959	17-5584	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR19394958	17-5586	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394957	17-5588	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394956	17-5589	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394955	17-5600	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394954	17-5606	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394953	17-5608	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394951	17-5610	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394950	17-5612	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394949	17-5615	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394948	17-5617	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394947	17-5618	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR19394946	17-5619	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394945	17-5620	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394944	17-5621	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394943	17-5629	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394942	17-5630	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394940	17-5635	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394939	17-5645	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394938	17-5646	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394937	17-5648	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394936	17-5650	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394935	17-5652	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394934	17-5653	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394933	17-5655	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR14863464	17-6354	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863463	17-6652	Ireland	Europe	UNK	UNK	13	RF-1
SRA Database	SRR14863461	17-6654	Ireland	Europe	UNK	UNK	11	RF-2
SRA Database	SRR14863460	17-670	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863459	17-6751	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863458	17-7446	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863457	17-7447	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863456	17-7448	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863455	17-7451	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863454	17-7457	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR19394932	18-6537	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394931	18-6540	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR19394929	18-6544	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394928	18-6545	Ireland	Europe	7	RF-1	14	RF-2
SRA Database	SRR19394927	18-6549	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR19394926	18-6552	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19394925	18-6555	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394924	18-6556	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394923	18-6560	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394922	18-6573	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394921	18-6574	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394920	18-6580	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394918	18-6587	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394917	18-6588	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394916	18-6590	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394915	18-6599	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395013	18-6602	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395012	18-6608	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395011	18-6611	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR19395010	18-6612	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395009	18-6613	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395008	18-6614	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395006	18-6615	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395005	18-6616	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395004	18-6618	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395003	18-6619	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395002	18-6622	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395001	18-6623	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395000	18-6624	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR19394999	18-6625	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394998	18-6626	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394997	18-6627	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394995	18-6628	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394994	18-6629	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394993	18-6630	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394992	18-6631	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394991	18-6632	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394989	18-6633	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394914	18-6634	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394913	18-6635	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394912	18-6636	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394911	18-6637	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394909	18-6638	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394908	18-6639	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394907	18-6640	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394906	18-6641	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394905	18-6642	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394904	18-6643	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR19394903	18-6645	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394902	18-6646	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR18116662	18b	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR18116661	18c	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR14863453	19-10903	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863452	19-10905	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863450	19-11225	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863449	19-11568	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863448	19-11570	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863447	19-2458b	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863446	19-2698	Ireland	Europe	UNK	UNK	11	RF-2
SRA Database	SRR14863445	19-2811	Ireland	Europe	11	RF-2	12	RF-3
SRA Database	SRR14863444	19-2812	Ireland	Europe	11	RF-2	11	RF-2
SRA Database	SRR14863443	19-3268	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394901	19-3336	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863442	19-3343	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863441	19-3474	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863439	19-3640	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863438	19-3680	Ireland	Europe	14	RF-2	10	RF-1
SRA Database	SRR14863437	19-3829	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863436	19-3830	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863435	19-3831	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863434	19-4608	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863433	19-4819	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863432	19-4820	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863431	19-4934	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863430	19-4935	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863428	19-5000	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863427	19-5575	Ireland	Europe	7	RF-1	10	RF-1

SRA Database	SRR14863426	19-5811	Ireland	Europe	9	RF-3	12	RF-3
SRA Database	SRR14863425	19-5812	Ireland	Europe	9	RF-3	12	RF-3
SRA Database	SRR14863424	19-5813	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394900	19-6407	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863423	19-6408	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863422	19-6576	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863421	19-6663	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863420	19-6893	Ireland	Europe	UNK	UNK	10	RF-1
SRA Database	SRR14863419	19-7585	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863417	19-7588	Ireland	Europe	UNK	UNK	9	RF-3
SRA Database	SRR14863416	19-7590	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395034	19-7591	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863415	19-7592	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395033	19-7601	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395032	19-7608	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395031	19-7611b	Ireland	Europe	9	RF-3	11	RF-2
SRA Database	SRR19395030	19-7613	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395029	19-7615	Ireland	Europe	9	RF-3	10	RF-1
SRA Database	SRR19395028	19-7617	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395027	19-7618	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395026	19-7619	Ireland	Europe	UNK	UNK	12	RF-3
SRA Database	SRR19395025	19-7620	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395023	19-7621	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395022	19-7622	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19395021	19-7623	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395020	19-7624	Ireland	Europe	UNK	UNK	11	RF-2
SRA Database	SRR19395019	19-7626	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395018	19-7628	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863414	19-7774	Ireland	Europe	7	RF-1	11	RF-2

SRA Database	SRR14863413	19-7917	Ireland	Europe	10	RF-1	11	RF-2
SRA Database	SRR14863412	19-7964	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863411	19-7966	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR14863410	19-7984	Ireland	Europe	9	RF-3	UNK	UNK
SRA Database	SRR14863409	19-7985	Ireland	Europe	8	RF-2	11	RF-2
SRA Database	SRR14863408	19-8272	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863406	19-8717	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863405	19-8936	Ireland	Europe	8	RF-2	10	RF-1
SRA Database	SRR18116656	21b	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR18116655	21c	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR18116654	22b	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR18116707	22c	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR12593858	6611_INMV1	Argentina	South America	7	RF-1	12	RF-3
SRA Database	SRR1793661	A1011	Canada	North America	10	RF-1	13	RF-1
SRA Database	SRR1793662	A1012	Canada	North America	UNK	UNK	17	RF-2
SRA Database	SRR1793663	A1013	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793664	A1014	Canada	North America	13	RF-1	9	RF-3
SRA Database	SRR1793665	A1016	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793666	A1018	Canada	North America	7	RF-1	9	RF-3
SRA Database	SRR1793667	A1019	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793668	A1020	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793669	A1021	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050003	A1023	Canada	North America	17	RF-2	10	RF-1
SRA Database	SRR3050004	A1028	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793672	A1034	Canada	North America	9	RF-3	11	RF-2
SRA Database	SRR1793673	A1035	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793674	A1037	Canada	North America	7	RF-1	9	RF-3
SRA Database	SRR1793675	A1038	Canada	North America	UNK	UNK	13	RF-1

SRA Database	SRR1793676	A1047	Canada	North America	9	RF-3	11	RF-2
SRA Database	SRR1793677	A1052	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793678	A1053	Canada	North America	9	RF-3	10	RF-1
SRA Database	SRR1793679	A1054	Canada	North America	7	RF-1	12	RF-3
SRA Database	SRR1793680	A1057	Canada	North America	UNK	UNK	13	RF-1
SRA Database	SRR1793681	A1060	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793683	A1067	Canada	North America	UNK	UNK	14	RF-2
SRA Database	SRR3050015	A1071	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793684	A1072	Canada	North America	11	RF-2	10	RF-1
SRA Database	SRR1793685	A1075	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793686	A1076	Canada	North America	16	RF-1	11	RF-2
SRA Database	SRR1793687	A1082	Canada	North America	UNK	UNK	14	RF-2
SRA Database	SRR1793688	A1085	Canada	North America	15	RF-3	10	RF-1
SRA Database	SRR1793689	A1092	Canada	North America	7	RF-1	13	RF-1
SRA Database	SRR1793690	A1095	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050026	A1099	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793691	A1102	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793693	A1105	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793696	A1110	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050037	A1112	Canada	North America	7	RF-1	15	RF-3
SRA Database	SRR1793697	A1113	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR3050048	A1114	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR1793698	A1124	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR1793699	A1125	Canada	North America	UNK	UNK	13	RF-1
SRA Database	SRR1793700	A1127	Canada	North America	UNK	UNK	8	RF-2
SRA Database	SRR3050059	A1129	Canada	North America	UNK	UNK	11	RF-2

SRA Database	SRR1793701	A1130	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050068	A1131	Canada	North America	10	RF-1	9	RF-3
SRA Database	SRR1793702	A1132	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793703	A1135	Canada	North America	13	RF-1	10	RF-1
SRA Database	SRR1793704	A1136	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR1793705	A1137	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793706	A1138	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793707	A1139	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR1793708	A1146	Canada	North America	9	RF-3	10	RF-1
SRA Database	SRR1793709	A1148	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793710	A1149	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR1793711	A1150	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR1793712	A1152	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793713	A1154	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793714	A1155	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR1793715	A1157	Canada	North America	13	RF-1	13	RF-1
SRA Database	SRR1793716	A1158	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793717	A1159	Canada	North America	7	RF-1	9	RF-3
SRA Database	SRR1793718	A1161	Canada	North America	14	RF-2	9	RF-3
SRA Database	SRR1793719	A1164	Canada	North America	7	RF-1	14	RF-2
SRA Database	SRR1793720	A1165	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793721	A1167	Canada	North America	7	RF-1	12	RF-3
SRA Database	SRR1793722	A1172	Canada	North America	UNK	UNK	13	RF-1
SRA Database	SRR1793723	A1175	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793724	A1179	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR1793725	A1180	Canada	North America	UNK	UNK	10	RF-1

SRA Database	SRR1793726	A1182	Canada	North America	10	RF-1	14	RF-2
SRA Database	SRR3050069	A1193	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793727	A1194	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793728	A1195	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793729	A1196	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793730	A1200	Canada	North America	12	RF-3	13	RF-1
SRA Database	SRR3050070	A1202	Canada	North America	7	RF-1	UNK	UNK
SRA Database	SRR1793731	A1203	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793733	A1214	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR1793734	A1216	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050005	A1223	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793737	A1229	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR1793738	A1252	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793740	A1255	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050006	A1279	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793741	A1285	Canada	North America	12	RF-3	14	RF-2
SRA Database	SRR1793742	A1286	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR1793743	A1288	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050007	A1291	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793744	A1293	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793745	A1310	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793746	A1312	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR1793747	A1320	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR1793749	A1326	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR1793750	A1331	Canada	North America	12	RF-3	9	RF-3
SRA Database	SRR1793752	A1334	Canada	North America	UNK	UNK	13	RF-1

SRA Database	SRR1793753	A1335	Canada	North America	8	RF-2	10	RF-1
SRA Database	SRR1793754	A1339	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793755	A1341	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR1793757	A1345	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR1793758	A1354	Canada	North America	UNK	UNK	14	RF-2
SRA Database	SRR1793759	A1356	Canada	North America	7	RF-1	8	RF-2
SRA Database	SRR1793760	A1357	Canada	North America	7	RF-1	12	RF-3
SRA Database	SRR1793761	A1362	Canada	North America	7	RF-1	13	RF-1
SRA Database	SRR1793762	A1364	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR1793765	A1370	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR1793766	A1371	Canada	North America	13	RF-1	9	RF-3
SRA Database	SRR1793767	A1374	Canada	North America	UNK	UNK	13	RF-1
SRA Database	SRR1793769	A1376	Canada	North America	7	RF-1	12	RF-3
SRA Database	SRR1793770	A1377	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR1793771	A1378	Canada	North America	8	RF-2	UNK	UNK
SRA Database	SRR1793772	A1380	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050008	A1512	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050009	A1514	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR3050010	A1522	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050011	A1525	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050012	A1526	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050013	A1527	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050014	A1529	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050016	A1530	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050017	A1534	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050018	A1536	Canada	North America	7	RF-1	10	RF-1

SRA Database	SRR3050019	A1538	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR3050020	A1545	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050021	A1546	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050022	A1547	Canada	North America	10	RF-1	10	RF-1
SRA Database	SRR3050023	A1548	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050024	A1590	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050025	A1622	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050027	A1627	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR3050028	A1629	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050029	A1635	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR3050030	A1636	Canada	North America	UNK	UNK	UNK	UNK
SRA Database	SRR3050031	A1637	Canada	North America	10	RF-1	12	RF-3
SRA Database	SRR3050032	A1638	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050033	A1639	Canada	North America	11	RF-2	12	RF-3
SRA Database	SRR3050034	A1640	Canada	North America	10	RF-1	12	RF-3
SRA Database	SRR3050035	A1641	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR3050036	A1643	Canada	North America	13	RF-1	9	RF-3
SRA Database	SRR3050038	A1644	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050039	A1645	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050040	A1646	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050041	A1647	Canada	North America	10	RF-1	11	RF-2
SRA Database	SRR3050042	A1648	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050043	A1649	Canada	North America	UNK	UNK	9	RF-3
SRA Database	SRR3050044	A1650	Canada	North America	10	RF-1	10	RF-1
SRA Database	SRR3050045	A1652	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050046	A1653	Canada	North America	UNK	UNK	12	RF-3

SRA Database	SRR3050047	A1654	Canada	North America	7	RF-1	12	RF-3
SRA Database	SRR3050049	A1655	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR3050050	A1656	Canada	North America	7	RF-1	9	RF-3
SRA Database	SRR3050051	A1659	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR3050052	A1661	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050053	A1664	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050054	A1667	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050055	A1668	Canada	North America	11	RF-2	10	RF-1
SRA Database	SRR3050056	A1669	Canada	North America	UNK	UNK	11	RF-2
SRA Database	SRR3050057	A1672	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050058	A1681	Canada	North America	8	RF-2	12	RF-3
SRA Database	SRR3050060	A1682	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR3050061	A1683	Canada	North America	7	RF-1	9	RF-3
SRA Database	SRR3050062	A1684	Canada	North America	UNK	UNK	12	RF-3
SRA Database	SRR3050063	A1685	Canada	North America	7	RF-1	9	RF-3
SRA Database	SRR3050064	A1686	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR3050065	A1NL1	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050066	A1NL2	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR3050067	A1NL3	Canada	North America	UNK	UNK	10	RF-1
SRA Database	SRR14863400	AFBI1	Ireland	Europe	9	RF-3	13	RF-1
SRA Database	SRR14863404	AFBI10	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863403	AFBI11	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863402	AFBI13	Ireland	Europe	8	RF-2	UNK	UNK
SRA Database	SRR14863401	AFBI14	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863399	AFBI2	Ireland	Europe	UNK	UNK	9	RF-3
SRA Database	SRR14863398	AFBI3	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863397	AFBI4	Ireland	Europe	7	RF-1	11	RF-2

SRA Database	SRR14863395	AFBI5	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863394	AFBI6	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863393	AFBI7	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863392	AFBI8	Ireland	Europe	10	RF-1	12	RF-3
SRA Database	SRR19395017	AFBI9	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR13195562	AFIF			UNK	UNK	13	RF-1
SRA Database	SRR19395016	c1	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394985	c10	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394984	c11	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR19394983	c12	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394982	c13	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394981	c14	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394980	c15	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394979	c16	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863391	c17	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863390	c18	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394977	c19	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19395015	c2	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863389	c20	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863388	c21	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863387	c22	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863386	c23	Ireland	Europe	7	RF-1	13	RF-1
SRA Database	SRR14863384	c24	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863383	c25	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863382	c26	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863381	c28	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863380	c29	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19395014	c3	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863379	c30	Ireland	Europe	7	RF-1	UNK	UNK

SRA Database	SRR14863378	c31	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863377	c32	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863376	c33	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863375	c34	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19394976	c35	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19394975	c36	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394974	c37	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394973	c38	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19394972	c39	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR14863369	c4	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863373	c40	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863372	c41	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR14863371	c42	Ireland	Europe	7	RF-1	9	RF-3
SRA Database	SRR14863370	c43	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR14863368	c5	Ireland	Europe	7	RF-1	10	RF-1
SRA Database	SRR19394988	c6	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394987	c7	Ireland	Europe	7	RF-1	UNK	UNK
SRA Database	SRR19394986	c8	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR14863367	c9	Ireland	Europe	7	RF-1	12	RF-3
SRA Database	SRR19394971	CIT	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR19394970	CITP	Ireland	Europe	7	RF-1	11	RF-2
SRA Database	SRR5413272	FDAARGOS_305	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR15101044	GT-9	India	Asia	7	RF-1	11	RF-2
SRA Database	SRR14335547	MAP 001	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335546	MAP 003	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335535	MAP 017	USA	North America	10	RF-1	9	RF-3
SRA Database	SRR14335524	MAP 071	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR14335513	MAP 120	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335502	MAP 122	USA	North America	7	RF-1	10	RF-1

SRA Database	SRR14335491	MAP 123	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335480	MAP 127	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335472	MAP 128	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR14335471	MAP 130	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335545	MAP 131	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335544	MAP 132	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335543	MAP 133	USA	North America	UNK	UNK	13	RF-1
SRA Database	SRR14335542	MAP 134	USA	North America	UNK	UNK	14	RF-2
SRA Database	SRR14335541	MAP 135	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335540	MAP 136	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335539	MAP 137	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335538	MAP 145	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335537	MAP 146	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335536	MAP 147	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335534	MAP 148	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335533	MAP 149	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335532	MAP 172	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335531	MAP 173	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335530	MAP 174	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335529	MAP 175	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335528	MAP 176	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335527	MAP 177	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335526	MAP 178	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335525	MAP 186	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR14335523	MAP 199	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335522	MAP 237	USA	North America	7	RF-1	10	RF-1

SRA Database	SRR14335521	MAP 238	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335520	MAP 239	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335519	MAP 240	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335518	MAP 313	USA	North America	14	RF-2	13	RF-1
SRA Database	SRR14335517	MAP 321	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335516	MAP 331	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335515	MAP 332	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335514	MAP 333	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335512	MAP 335	USA	North America	UNK	UNK	12	RF-3
SRA Database	SRR14335511	MAP 336	USA	North America	UNK	UNK	13	RF-1
SRA Database	SRR14335510	MAP 337	USA	North America	UNK	UNK	12	RF-3
SRA Database	SRR14335509	MAP 339	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335508	MAP 340	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335507	MAP 341	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR14335506	MAP 353	USA	North America	11	RF-2	11	RF-2
SRA Database	SRR14335505	MAP 355	USA	North America	11	RF-2	12	RF-3
SRA Database	SRR14335504	MAP 356	USA	North America	11	RF-2	11	RF-2
SRA Database	SRR14335503	MAP 357	USA	North America	10	RF-1	9	RF-3
SRA Database	SRR14335501	MAP 358	USA	North America	11	RF-2	12	RF-3
SRA Database	SRR14335500	MAP 388	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335499	MAP 425	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335498	MAP 435	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335497	MAP 460	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335496	MAP 469	USA	North America	UNK	UNK	13	RF-1
SRA Database	SRR14335495	MAP 470	USA	North America	UNK	UNK	13	RF-1
SRA Database	SRR14335494	MAP 486	USA	North America	7	RF-1	10	RF-1

SRA Database	SRR14335493	MAP 494	USA	North America	UNK	UNK	13	RF-1
SRA Database	SRR14335492	MAP 519	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335490	MAP 520	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335489	MAP 521	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335488	MAP 522	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335487	MAP 523	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335486	MAP 562	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335485	MAP 565	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335484	MAP 573	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335483	MAP 576	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335482	MAP 579	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335481	MAP 585	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335479	MAP 603	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR14335478	MAP 617	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335477	MAP 618	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335476	MAP 619	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335475	MAP 620	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335474	MAP 621	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR14335473	MAP 622	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR12237016	Map 907-K32	Argentina	South America	7	RF-1	10	RF-1
SRA Database	SRR12237017	Map L80	Argentina	South America	7	RF-1	11	RF-2
SRA Database	SRR22426156	Map LN20	Canada	North America	7	RF-1	10	RF-1
SRA Database	SRR5320599	MAP NL 89C	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR5320598	MAP NL 93B	Canada	North America	11	RF-2	11	RF-2
SRA Database	SRR5320597	MAP NL 95A	Canada	North America	10	RF-1	11	RF-2
SRA Database	SRR5320596	MAP NL 95B	Canada	North America	UNK	UNK	11	RF-2

SRA Database	SRR5320595	MAP NL 95E	Canada	North America	7	RF-1	11	RF-2
SRA Database	SRR5320594	MAP NL 96E	Canada	North America	12	RF-3	10	RF-1
SRA Database	SRR201793	MAP_CLIJ361	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR201790	MAP_CLIJ623	Australia	Oceania	19	RF-1	10	RF-1
SRA Database	SRR201791	MAP_CLIJ644	Australia	Oceania	7	RF-1	19	RF-1
SRA Database	SRR13395951	MAP004	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395950	MAP007	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395839	MAP008	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395728	MAP009	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396016	MAP010	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR13396006	MAP011	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395995	MAP012	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395984	MAP014	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR13395973	MAP015	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395962	MAP016	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395949	MAP018	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395938	MAP019	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395927	MAP020	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395916	MAP021	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395905	MAP022	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR13395894	MAP023	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395883	MAP025	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395872	MAP028	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395861	MAP029	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395850	MAP030	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395838	MAP032	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395827	MAP033	USA	North America	7	RF-1	10	RF-1

SRA Database	SRR13395816	MAP050	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395805	MAP051	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395794	MAP054	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395783	MAP055	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395772	MAP058	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395761	MAP059	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395750	MAP060	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395739	MAP061	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395727	MAP066	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396033	MAP082	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396024	MAP083	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396023	MAP086	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13396022	MAP087	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396021	MAP091	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396020	MAP092	USA	North America	8	RF-2	11	RF-2
SRA Database	SRR13396019	MAP094	USA	North America	UNK	UNK	11	RF-2
SRA Database	SRR13396018	MAP097	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR17460194	Map1	France	Europe	7	RF-1	9	RF-3
SRA Database	SRR13396017	MAP102	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839016	MAP-102	Australia	Oceania	7	RF-1	UNK	UNK
SRA Database	SRR11839015	MAP-106	Australia	Oceania	13	RF-1	11	RF-2
SRA Database	SRR13396015	MAP107	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396014	MAP108	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839093	MAP-108	Australia	Oceania	UNK	UNK	UNK	UNK
SRA Database	SRR13395717	MAP109	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396013	MAP110	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13396012	MAP111	USA	North America	7	RF-1	11	RF-2

SRA Database	SRR13396011	MAP112	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839082	MAP-112	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR13396010	MAP113	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839071	MAP-113	Australia	Oceania	13	RF-1	13	RF-1
SRA Database	SRR11839060	MAP-114	Australia	Oceania	12	RF-3	12	RF-3
SRA Database	SRR11839027	MAP-118	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13396009	MAP119	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839003	MAP-120	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13396008	MAP121	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838992	MAP-121	France	Europe	UNK	UNK	9	RF-3
SRA Database	SRR11838981	MAP-122	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11838970	MAP-123	Australia	Oceania	11	RF-2	9	RF-3
SRA Database	SRR13396007	MAP124	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838959	MAP-125	Australia	Oceania	UNK	UNK	12	RF-3
SRA Database	SRR13396005	MAP126	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838948	MAP-127	Australia	Oceania	14	RF-2	11	RF-2
SRA Database	SRR11838937	MAP-128	Australia	Oceania	7	RF-1	13	RF-1
SRA Database	SRR11838926	MAP-129	Australia	Oceania	13	RF-1	11	RF-2
SRA Database	SRR11839145	MAP-131	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11839133	MAP-132	Australia	Oceania	18	RF-3	9	RF-3
SRA Database	SRR11839122	MAP-134	Australia	Oceania	7	RF-1	13	RF-1
SRA Database	SRR11839111	MAP-135	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR11839100	MAP-138	Australia	Oceania	15	RF-3	13	RF-1
SRA Database	SRR13396004	MAP139	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR17460193	Map139	France	Europe	UNK	UNK	12	RF-3
SRA Database	SRR11839099	MAP-139	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR17460192	Map14	France	Europe	7	RF-1	10	RF-1
SRA Database	SRR17460191	Map140	France	Europe	7	RF-1	11	RF-2
SRA Database	SRR11839098	MAP-140	Australia	Oceania	7	RF-1	10	RF-1

SRA Database	SRR11839097	MAP-141	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13396003	MAP142	USA	North America	8	RF-2	10	RF-1
SRA Database	SRR11839096	MAP-142	Australia	Oceania	7	RF-1	13	RF-1
SRA Database	SRR13396002	MAP143	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR11839095	MAP-143	Australia	Oceania	11	RF-2	11	RF-2
SRA Database	SRR13396001	MAP144	USA	North America	7	RF-1	13	RF-1
SRA Database	SRR11839094	MAP-144	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11839092	MAP-147	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11839091	MAP-148	Australia	Oceania	UNK	UNK	13	RF-1
SRA Database	SRR11839090	MAP-149	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR13396000	MAP150	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839089	MAP-150	Australia	Oceania	10	RF-1	9	RF-3
SRA Database	SRR13395999	MAP151	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395998	MAP152	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839088	MAP-152	Australia	Oceania	14	RF-2	10	RF-1
SRA Database	SRR11839087	MAP-153	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395997	MAP154	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839086	MAP-154	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR13395996	MAP155	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839085	MAP-155	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395994	MAP156	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839084	MAP-156	Australia	Oceania	12	RF-3	9	RF-3
SRA Database	SRR11839083	MAP-157	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395993	MAP158	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839081	MAP-158	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395992	MAP159	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR17460190	Map159	France	Europe	7	RF-1	11	RF-2
SRA Database	SRR11839080	MAP-159	Australia	Oceania	7	RF-1	13	RF-1

SRA Database	SRR13395991	MAP160	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839079	MAP-160	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395990	MAP161	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839078	MAP-161	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395989	MAP162	USA	North America	UNK	UNK	9	RF-3
SRA Database	SRR11839077	MAP-162	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR13395988	MAP163	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839076	MAP-163	Australia	Oceania	13	RF-1	11	RF-2
SRA Database	SRR11839075	MAP-164	Australia	Oceania	10	RF-1	10	RF-1
SRA Database	SRR13395987	MAP165	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839074	MAP-165	Australia	Oceania	7	RF-1	9	RF-3
SRA Database	SRR13395986	MAP166	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839073	MAP-166	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395985	MAP167	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839072	MAP-167	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395983	MAP168	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395982	MAP169	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839070	MAP-169	Australia	Oceania	11	RF-2	9	RF-3
SRA Database	SRR11839054	MAP-170	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11839069	MAP-170	Australia	Oceania	11	RF-2	9	RF-3
SRA Database	SRR13395981	MAP171	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839053	MAP-171	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR11839068	MAP-171	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR11839052	MAP-172	Australia	Oceania	17	RF-2	UNK	UNK
SRA Database	SRR11839067	MAP-172	Australia	Oceania	12	RF-3	9	RF-3
SRA Database	SRR11839051	MAP-174	France	Europe	7	RF-1	9	RF-3
SRA Database	SRR11839066	MAP-174	France	Europe	7	RF-1	11	RF-2
SRA Database	SRR11839050	MAP-175	France	Europe	7	RF-1	11	RF-2

SRA Database	SRR11839065	MAP-175	France	Europe	7	RF-1	11	RF-2
SRA Database	SRR11839048	MAP-176	France	Europe	7	RF-1	12	RF-3
SRA Database	SRR11839064	MAP-176	France	Europe	7	RF-1	9	RF-3
SRA Database	SRR13395980	MAP179	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395979	MAP180	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395978	MAP181	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395977	MAP182	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395976	MAP183	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395975	MAP184	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395974	MAP185	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395972	MAP187	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR17460189	Map187B	France	Europe	7	RF-1	11	RF-2
SRA Database	SRR13395971	MAP189	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR17460188	Map19	France	Europe	UNK	UNK	11	RF-2
SRA Database	SRR13395970	MAP190	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR13395969	MAP191	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395968	MAP193	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395967	MAP194	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395966	MAP195	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395965	MAP196	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395964	MAP197	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395963	MAP198	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395961	MAP200	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395960	MAP201	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395959	MAP202	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839047	MAP-202	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11839063	MAP-202	Australia	Oceania	7	RF-1	12	RF-3

SRA Database	SRR11839046	MAP-203	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11839062	MAP-203	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395958	MAP204	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR11839045	MAP-204	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11839061	MAP-204	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395957	MAP205	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR11839044	MAP-205	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR11839059	MAP-205	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395956	MAP206	USA	North America	8	RF-2	11	RF-2
SRA Database	SRR11839043	MAP-206	Australia	Oceania	7	RF-1	13	RF-1
SRA Database	SRR11839058	MAP-206	Australia	Oceania	16	RF-1	10	RF-1
SRA Database	SRR13395955	MAP207	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839042	MAP-207	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR11839057	MAP-207	Australia	Oceania	10	RF-1	9	RF-3
SRA Database	SRR13395954	MAP208	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839041	MAP-208	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11839056	MAP-208	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR11839040	MAP-209	Australia	Oceania	7	RF-1	15	RF-3
SRA Database	SRR11839055	MAP-209	Australia	Oceania	13	RF-1	11	RF-2
SRA Database	SRR13395953	MAP210	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395952	MAP211	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395948	MAP213	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395947	MAP214	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395946	MAP215	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395945	MAP216	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395944	MAP217	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395943	MAP220	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395942	MAP221	USA	North America	7	RF-1	10	RF-1

SRA Database	SRR13395941	MAP222	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395940	MAP223	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395939	MAP224	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395937	MAP225	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839039	MAP-225	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395936	MAP226	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839037	MAP-226	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR13395935	MAP227	USA	North America	7	RF-1	13	RF-1
SRA Database	SRR11839036	MAP-227	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395934	MAP228	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839035	MAP-228	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11839034	MAP-229	Australia	Oceania	15	RF-3	10	RF-1
SRA Database	SRR13395933	MAP230	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839033	MAP-230	Australia	Oceania	16	RF-1	13	RF-1
SRA Database	SRR11839032	MAP-233	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR11839031	MAP-235	Australia	Oceania	10	RF-1	10	RF-1
SRA Database	SRR13395932	MAP241	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839030	MAP-241	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR11839029	MAP-242	Australia	Oceania	13	RF-1	12	RF-3
SRA Database	SRR11839028	MAP-243	Australia	Oceania	UNK	UNK	12	RF-3
SRA Database	SRR11839026	MAP-244	Australia	Oceania	13	RF-1	11	RF-2
SRA Database	SRR11839025	MAP-245	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR11839024	MAP-246	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR11839023	MAP-247	Australia	Oceania	13	RF-1	12	RF-3
SRA Database	SRR11839022	MAP-249	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR11839021	MAP-250	Australia	Oceania	13	RF-1	12	RF-3
SRA Database	SRR11839020	MAP-251	Australia	Oceania	15	RF-3	11	RF-2
SRA Database	SRR11839019	MAP-252	Australia	Oceania	12	RF-3	10	RF-1

SRA Database	SRR11839018	MAP-253	Australia	Oceania	14	RF-2	10	RF-1
SRA Database	SRR11839017	MAP-254	Australia	Oceania	18	RF-3	11	RF-2
SRA Database	SRR11839013	MAP-271	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR11839012	MAP-273	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR11839011	MAP-274	Australia	Oceania	9	RF-3	12	RF-3
SRA Database	SRR13395931	MAP283	USA	North America	12	RF-3	10	RF-1
SRA Database	SRR13395930	MAP284	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395929	MAP285	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395928	MAP286	USA	North America	12	RF-3	12	RF-3
SRA Database	SRR13395926	MAP287	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR13395925	MAP288	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR13395924	MAP289	USA	North America	13	RF-1	10	RF-1
SRA Database	SRR13395923	MAP290	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395922	MAP291	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395921	MAP292	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839010	MAP-304	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11839009	MAP-305	Australia	Oceania	7	RF-1	13	RF-1
SRA Database	SRR11839008	MAP-306	Australia	Oceania	9	RF-3	18	RF-3
SRA Database	SRR13395920	MAP307	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR13395919	MAP308	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395918	MAP312	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR13395917	MAP319	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395915	MAP320	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839007	MAP-320	Australia	Oceania	15	RF-3	12	RF-3
SRA Database	SRR11839006	MAP-322	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR11839005	MAP-323	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395914	MAP324	USA	North America	7	RF-1	13	RF-1
SRA Database	SRR11839004	MAP-324	Australia	Oceania	9	RF-3	14	RF-2

SRA Database	SRR13395913	MAP327	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839002	MAP-327	Australia	Oceania	9	RF-3	13	RF-1
SRA Database	SRR11839001	MAP-328	Australia	Oceania	9	RF-3	16	RF-1
SRA Database	SRR13395912	MAP329	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR11839000	MAP-330	Australia	Oceania	9	RF-3	13	RF-1
SRA Database	SRR11838999	MAP-332	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR11838998	MAP-333	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395911	MAP334	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395910	MAP338	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11838997	MAP-339	Australia	Oceania	8	RF-2	11	RF-2
SRA Database	SRR11838996	MAP-340	Australia	Oceania	8	RF-2	12	RF-3
SRA Database	SRR11838995	MAP-341	Australia	Oceania	8	RF-2	10	RF-1
SRA Database	SRR13395909	MAP342	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838994	MAP-342	Australia	Oceania	8	RF-2	12	RF-3
SRA Database	SRR13395908	MAP343	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR13395907	MAP344	USA	North America	UNK	UNK	9	RF-3
SRA Database	SRR13395906	MAP345	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR13395904	MAP346	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR11838993	MAP-346	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR13395903	MAP347	USA	North America	21	RF-3	10	RF-1
SRA Database	SRR11838991	MAP-347	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR11838990	MAP-348	Australia	Oceania	8	RF-2	10	RF-1
SRA Database	SRR13395902	MAP349	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838989	MAP-349	Australia	Oceania	9	RF-3	14	RF-2
SRA Database	SRR13395901	MAP350	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838988	MAP-350	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395900	MAP351	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838987	MAP-351	Australia	Oceania	9	RF-3	12	RF-3

SRA Database	SRR13395899	MAP352	USA	North America	UNK	UNK	10	RF-1
SRA Database	SRR11838986	MAP-353	Australia	Oceania	17	RF-2	11	RF-2
SRA Database	SRR13395898	MAP359	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395897	MAP362	USA	North America	UNK	UNK	9	RF-3
SRA Database	SRR11838985	MAP-372	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395896	MAP374	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395895	MAP376	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395893	MAP378	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR17460187	Map38	France	Europe	7	RF-1	UNK	UNK
SRA Database	SRR13395892	MAP380	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395891	MAP381	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395890	MAP383	USA	North America	11	RF-2	13	RF-1
SRA Database	SRR13395889	MAP384	USA	North America	UNK	UNK	11	RF-2
SRA Database	SRR13395888	MAP387	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395887	MAP391	USA	North America	14	RF-2	11	RF-2
SRA Database	SRR13395886	MAP392	USA	North America	UNK	UNK	11	RF-2
SRA Database	SRR13395885	MAP393	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395884	MAP394	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395882	MAP396	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395881	MAP398	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR17460186	Map40	France	Europe	7	RF-1	12	RF-3
SRA Database	SRR13395880	MAP401	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395879	MAP402	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838984	MAP-403	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR13395878	MAP404	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838983	MAP-404	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR13395877	MAP407	USA	North America	7	RF-1	9	RF-3

SRA Database	SRR11838982	MAP-407	Australia	Oceania	7	RF-1	14	RF-2
SRA Database	SRR13395876	MAP408	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838980	MAP-408	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395875	MAP409	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838979	MAP-409	Australia	Oceania	14	RF-2	10	RF-1
SRA Database	SRR11838978	MAP-411	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR11838977	MAP-413	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR11838976	MAP-414	Australia	Oceania	9	RF-3	14	RF-2
SRA Database	SRR13395874	MAP415	USA	North America	8	RF-2	9	RF-3
SRA Database	SRR11838975	MAP-415	Australia	Oceania	UNK	UNK	9	RF-3
SRA Database	SRR11838974	MAP-416	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11838973	MAP-417	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR13395873	MAP418	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838972	MAP-418	Australia	Oceania	7	RF-1	9	RF-3
SRA Database	SRR11838971	MAP-419	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR11838969	MAP-420	Australia	Oceania	7	RF-1	16	RF-1
SRA Database	SRR13395871	MAP422	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11838968	MAP-422	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395870	MAP423	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838967	MAP-423	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR13395869	MAP424	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838966	MAP-424	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11838965	MAP-425	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11838964	MAP-426	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11838963	MAP-428	Australia	Oceania	15	RF-3	10	RF-1
SRA Database	SRR13395868	MAP429	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838962	MAP-429	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11838961	MAP-431	Australia	Oceania	UNK	UNK	9	RF-3
SRA Database	SRR11838960	MAP-434	Australia	Oceania	18	RF-3	10	RF-1

SRA Database	SRR11838958	MAP-435	Australia	Oceania	UNK	UNK	13	RF-1
SRA Database	SRR13395867	MAP436	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838957	MAP-436	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR13395866	MAP437	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395865	MAP438	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838956	MAP-438	Australia	Oceania	7	RF-1	14	RF-2
SRA Database	SRR13395864	MAP439	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838955	MAP-439	Australia	Oceania	UNK	UNK	12	RF-3
SRA Database	SRR13395863	MAP440	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838954	MAP-440	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395862	MAP442	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838953	MAP-442	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395860	MAP443	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838952	MAP-443	Australia	Oceania	UNK	UNK	10	RF-1
SRA Database	SRR11838951	MAP-444	Australia	Oceania	UNK	UNK	9	RF-3
SRA Database	SRR11838950	MAP-445	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR13395859	MAP450	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395858	MAP452	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395857	MAP453	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395856	MAP454	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395855	MAP455	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395854	MAP456	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395853	MAP457	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395852	MAP458	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395851	MAP459	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395849	MAP461	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838949	MAP-461	Australia	Oceania	7	RF-1	13	RF-1

SRA Database	SRR13395848	MAP464	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395847	MAP465	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395846	MAP471	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395845	MAP473	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395844	MAP474	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395843	MAP475	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395842	MAP476	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395841	MAP477	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395840	MAP478	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395837	MAP479	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395836	MAP480	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395835	MAP481	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395834	MAP482	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395833	MAP483	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395832	MAP484	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395831	MAP485	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395830	MAP487	USA	North America	13	RF-1	10	RF-1
SRA Database	SRR13395829	MAP488	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395828	MAP489	USA	North America	7	RF-1	12	RF-3
SRA Database	SRR13395826	MAP490	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395825	MAP491	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395824	MAP492	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395823	MAP493	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395822	MAP495	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395821	MAP496	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395820	MAP497	USA	North America	7	RF-1	9	RF-3

SRA Database	SRR13395819	MAP498	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395818	MAP499	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395817	MAP500	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395815	MAP501	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395814	MAP502	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395813	MAP503	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395812	MAP504	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395811	MAP505	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395810	MAP506	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838947	MAP-506	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395809	MAP507	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395808	MAP508	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395807	MAP509	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838946	MAP-509	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395806	MAP510	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395804	MAP511	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395803	MAP512	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395802	MAP513	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838945	MAP-513	Australia	Oceania	7	RF-1	9	RF-3
SRA Database	SRR13395801	MAP514	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838944	MAP-514	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR13395800	MAP515	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838943	MAP-515	Australia	Oceania	10	RF-1	12	RF-3
SRA Database	SRR13395799	MAP516	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395798	MAP517	USA	North America	8	RF-2	10	RF-1
SRA Database	SRR11838942	MAP-517	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR13395797	MAP518	USA	North America	7	RF-1	9	RF-3

SRA Database	SRR11838941	MAP-520	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11838940	MAP-521	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395796	MAP524	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11838939	MAP-524	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395795	MAP525	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838938	MAP-525	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395793	MAP526	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838936	MAP-526	Australia	Oceania	UNK	UNK	12	RF-3
SRA Database	SRR13395792	MAP527	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838935	MAP-527	Australia	Oceania	12	RF-3	10	RF-1
SRA Database	SRR13395791	MAP528	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395790	MAP529	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838934	MAP-529	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395789	MAP531	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395788	MAP532	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395787	MAP533	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838933	MAP-533	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13395786	MAP534	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838932	MAP-534	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR11838931	MAP-535	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395785	MAP536	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838930	MAP-536	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395784	MAP537	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR13395782	MAP538	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838929	MAP-538	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395781	MAP539	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838928	MAP-539	Australia	Oceania	7	RF-1	9	RF-3
SRA Database	SRR11838927	MAP-540	Australia	Oceania	7	RF-1	13	RF-1

SRA Database	SRR13395780	MAP541	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838925	MAP-541	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395779	MAP542	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395778	MAP543	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395777	MAP544	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395776	MAP545	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395775	MAP546	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395774	MAP547	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395773	MAP548	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395771	MAP549	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395770	MAP550	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395769	MAP551	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838924	MAP-551	Australia	Oceania	UNK	UNK	9	RF-3
SRA Database	SRR13395768	MAP552	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838923	MAP-552	Australia	Oceania	17	RF-2	9	RF-3
SRA Database	SRR13395767	MAP553	USA	North America	9	RF-3	11	RF-2
SRA Database	SRR13395766	MAP554	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838922	MAP-554	Australia	Oceania	9	RF-3	11	RF-2
SRA Database	SRR13395765	MAP555	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838921	MAP-555	Australia	Oceania	7	RF-1	14	RF-2
SRA Database	SRR13395764	MAP556	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838920	MAP-556	Australia	Oceania	15	RF-3	9	RF-3
SRA Database	SRR13395763	MAP557	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838919	MAP-557	Australia	Oceania	16	RF-1	9	RF-3
SRA Database	SRR13395762	MAP558	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11838918	MAP-558	Australia	Oceania	8	RF-2	12	RF-3
SRA Database	SRR13395760	MAP559	USA	North America	7	RF-1	9	RF-3

SRA Database	SRR11839147	MAP-559	Australia	Oceania	8	RF-2	12	RF-3
SRA Database	SRR13395759	MAP560	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839146	MAP-560	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR11839144	MAP-561	Australia	Oceania	UNK	UNK	UNK	UNK
SRA Database	SRR11839143	MAP-562	Australia	Oceania	8	RF-2	12	RF-3
SRA Database	SRR13395758	MAP563	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395757	MAP564	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839142	MAP-564	Australia	Oceania	14	RF-2	12	RF-3
SRA Database	SRR13395756	MAP566	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395755	MAP567	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839141	MAP-567	Australia	Oceania	13	RF-1	12	RF-3
SRA Database	SRR13395754	MAP568	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839140	MAP-568	Australia	Oceania	14	RF-2	13	RF-1
SRA Database	SRR13395753	MAP569	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395752	MAP570	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839139	MAP-570	Australia	Oceania	13	RF-1	15	RF-3
SRA Database	SRR13395751	MAP571	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839138	MAP-571	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395749	MAP572	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395748	MAP574	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395747	MAP575	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395746	MAP577	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839137	MAP-577	Australia	Oceania	16	RF-1	11	RF-2
SRA Database	SRR13395745	MAP578	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839136	MAP-579	Australia	Oceania	14	RF-2	11	RF-2
SRA Database	SRR13395744	MAP580	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839135	MAP-580	Australia	Oceania	11	RF-2	11	RF-2
SRA Database	SRR13395743	MAP581	USA	North America	7	RF-1	9	RF-3

SRA Database	SRR11839132	MAP-581	Australia	Oceania	11	RF-2	11	RF-2
SRA Database	SRR13395742	MAP582	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395741	MAP583	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839131	MAP-583	Australia	Oceania	13	RF-1	10	RF-1
SRA Database	SRR13395740	MAP584	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839130	MAP-584	Australia	Oceania	12	RF-3	12	RF-3
SRA Database	SRR11839129	MAP-585	Australia	Oceania	16	RF-1	10	RF-1
SRA Database	SRR13395738	MAP586	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839128	MAP-586	Australia	Oceania	15	RF-3	10	RF-1
SRA Database	SRR13395737	MAP587	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839127	MAP-587	Australia	Oceania	15	RF-3	11	RF-2
SRA Database	SRR13395736	MAP588	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839126	MAP-588	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR11839125	MAP-589	Australia	Oceania	UNK	UNK	11	RF-2
SRA Database	SRR13395735	MAP590	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395734	MAP591	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839124	MAP-592	Australia	Oceania	18	RF-3	10	RF-1
SRA Database	SRR13395733	MAP593	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395732	MAP594	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839123	MAP-594	Australia	Oceania	11	RF-2	9	RF-3
SRA Database	SRR13395731	MAP595	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839121	MAP-595	Australia	Oceania	15	RF-3	11	RF-2
SRA Database	SRR13395730	MAP596	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839120	MAP-596	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13395729	MAP597	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839119	MAP-597	Australia	Oceania	11	RF-2	10	RF-1
SRA Database	SRR13395726	MAP598	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839118	MAP-598	Australia	Oceania	9	RF-3	12	RF-3

SRA Database	SRR13395725	MAP599	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839117	MAP-599	Australia	Oceania	7	RF-1	15	RF-3
SRA Database	SRR17460185	Map6	France	Europe	7	RF-1	13	RF-1
SRA Database	SRR13395724	MAP600	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395723	MAP601	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395722	MAP602	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839116	MAP-602	Australia	Oceania	8	RF-2	10	RF-1
SRA Database	SRR13395721	MAP604	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR13395720	MAP605	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395719	MAP606	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13395718	MAP607	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839115	MAP-607	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13395716	MAP608	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839114	MAP-608	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13396032	MAP609	USA	North America	7	RF-1	11	RF-2
SRA Database	SRR11839113	MAP-609	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR13396031	MAP610	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839112	MAP-610	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR13396030	MAP611	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839110	MAP-611	Australia	Oceania	7	RF-1	11	RF-2
SRA Database	SRR13396029	MAP612	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839109	MAP-612	Australia	Oceania	9	RF-3	13	RF-1
SRA Database	SRR13396028	MAP613	USA	North America	7	RF-1	10	RF-1
SRA Database	SRR11839108	MAP-613	Australia	Oceania	10	RF-1	14	RF-2
SRA Database	SRR13396027	MAP614	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13396026	MAP615	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR13396025	MAP616	USA	North America	7	RF-1	9	RF-3
SRA Database	SRR11839107	MAP-617	Australia	Oceania	15	RF-3	11	RF-2

SRA Database	SRR11839106	MAP-618	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR11839105	MAP-619	Australia	Oceania	7	RF-1	10	RF-1
SRA Database	SRR11839104	MAP-620	Australia	Oceania	9	RF-3	9	RF-3
SRA Database	SRR11839103	MAP-621	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR11839102	MAP-622	Australia	Oceania	9	RF-3	10	RF-1
SRA Database	SRR11839101	MAP-623	Australia	Oceania	7	RF-1	12	RF-3
SRA Database	SRR11802323	Mycobacterium avium subsp. paratuberculosis DSM 44135	Germany	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415108	PICSAR1	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415109	PICSAR10	France	Europe	9	RF-3	13	RF-1
SRA Database	ERR6415110	PICSAR100	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415111	PICSAR102	France	Europe	9	RF-3	11	RF-2
SRA Database	ERR6415112	PICSAR103	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415113	PICSAR104	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415114	PICSAR106	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415115	PICSAR107	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415116	PICSAR11	France	Europe	UNK	UNK	10	RF-1
SRA Database	ERR6415117	PICSAR110	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415118	PICSAR111	France	Europe	13	RF-1	13	RF-1
SRA Database	ERR6415119	PICSAR113	France	Europe	7	RF-1	13	RF-1
SRA Database	ERR6415288	PICSAR114	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415120	PICSAR117	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415121	PICSAR118	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415122	PICSAR119	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415123	PICSAR120	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415125	PICSAR124	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415124	PICSAR124B	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415126	PICSAR126	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415253	PICSAR127	France	Europe	12	RF-3	10	RF-1
SRA Database	ERR6415289	PICSAR128	France	Europe	7	RF-1	10	RF-1

SRA Database	ERR6415127	PICSAR13	France	Europe	8	RF-2	11	RF-2
SRA Database	ERR6415128	PICSAR131	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415129	PICSAR132	France	Europe	7	RF-1	13	RF-1
SRA Database	ERR6415130	PICSAR135	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415131	PICSAR136	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415240	PICSAR138	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415255	PICSAR139	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415132	PICSAR14	France	Europe	8	RF-2	10	RF-1
SRA Database	ERR6415133	PICSAR140	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415134	PICSAR141	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415135	PICSAR142	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415136	PICSAR143	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415249	PICSAR144	France	Europe	7	RF-1	13	RF-1
SRA Database	ERR6415137	PICSAR145	France	Europe	8	RF-2	14	RF-2
SRA Database	ERR6415256	PICSAR146	France	Europe	7	RF-1	UNK	UNK
SRA Database	ERR6415138	PICSAR147	France	Europe	7	RF-1	13	RF-1
SRA Database	ERR6415139	PICSAR148	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415290	PICSAR149	France	Europe	10	RF-1	11	RF-2
SRA Database	ERR6415140	PICSAR15	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415291	PICSAR153	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415141	PICSAR154	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415142	PICSAR157	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415143	PICSAR16	France	Europe	9	RF-3	10	RF-1
SRA Database	ERR6415144	PICSAR160	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415145	PICSAR162	France	Europe	8	RF-2	13	RF-1
SRA Database	ERR6415146	PICSAR164	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415148	PICSAR168	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415147	PICSAR168B	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415292	PICSAR17	France	Europe	8	RF-2	11	RF-2

SRA Database	ERR6415244	PICSAR171	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415149	PICSAR172	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415150	PICSAR178	France	Europe	8	RF-2	11	RF-2
SRA Database	ERR6415151	PICSAR179	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415152	PICSAR18	France	Europe	7	RF-1	UNK	UNK
SRA Database	ERR6415153	PICSAR180	France	Europe	UNK	UNK	11	RF-2
SRA Database	ERR6415154	PICSAR181	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415155	PICSAR182	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415156	PICSAR184	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415157	PICSAR189	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415158	PICSAR19	France	Europe	8	RF-2	14	RF-2
SRA Database	ERR6415248	PICSAR190	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415159	PICSAR191	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415160	PICSAR192	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415161	PICSAR196	France	Europe	8	RF-2	10	RF-1
SRA Database	ERR6415162	PICSAR197	France	Europe	8	RF-2	10	RF-1
SRA Database	ERR6415163	PICSAR198	France	Europe	7	RF-1	8	RF-2
SRA Database	ERR6415164	PICSAR2	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415293	PICSAR20	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415245	PICSAR200	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415165	PICSAR201	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415166	PICSAR202	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415263	PICSAR203	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415167	PICSAR204	France	Europe	8	RF-2	10	RF-1
SRA Database	ERR6415168	PICSAR205	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415169	PICSAR206	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415170	PICSAR208	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415171	PICSAR209	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415172	PICSAR21	France	Europe	7	RF-1	10	RF-1

SRA Database	ERR6415173	PICSAR210	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415174	PICSAR211	France	Europe	UNK	UNK	9	RF-3
SRA Database	ERR6415264	PICSAR212	France	Europe	8	RF-2	12	RF-3
SRA Database	ERR6415242	PICSAR213	France	Europe	8	RF-2	11	RF-2
SRA Database	ERR6415175	PICSAR214	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415176	PICSAR215	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415177	PICSAR216	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415178	PICSAR218	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415265	PICSAR219	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415179	PICSAR22	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415180	PICSAR222	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415266	PICSAR223	France	Europe	10	RF-1	12	RF-3
SRA Database	ERR6415181	PICSAR23	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415182	PICSAR232	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415183	PICSAR235	France	Europe	7	RF-1	UNK	UNK
SRA Database	ERR6415246	PICSAR236	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415184	PICSAR238	France	Europe	8	RF-2	11	RF-2
SRA Database	ERR6415185	PICSAR24	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415186	PICSAR240	France	Europe	11	RF-2	12	RF-3
SRA Database	ERR6415273	PICSAR245	France	Europe	8	RF-2	UNK	UNK
SRA Database	ERR6415187	PICSAR246	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415274	PICSAR247	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415188	PICSAR248	France	Europe	7	RF-1	14	RF-2
SRA Database	ERR6415189	PICSAR249	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415190	PICSAR25	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415191	PICSAR250	France	Europe	7	RF-1	8	RF-2
SRA Database	ERR6415294	PICSAR251	France	Europe	7	RF-1	8	RF-2
SRA Database	ERR6415192	PICSAR252	France	Europe	7	RF-1	8	RF-2
SRA Database	ERR6415193	PICSAR253	France	Europe	7	RF-1	10	RF-1

SRA Database	ERR6415194	PICSAR254	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415195	PICSAR255	France	Europe	7	RF-1	UNK	UNK
SRA Database	ERR6415275	PICSAR257	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415196	PICSAR26	France	Europe	9	RF-3	11	RF-2
SRA Database	ERR6415276	PICSAR27	France	Europe	9	RF-3	9	RF-3
SRA Database	ERR6415197	PICSAR28	France	Europe	10	RF-1	10	RF-1
SRA Database	ERR6415198	PICSAR29	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415199	PICSAR3	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415200	PICSAR30	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415247	PICSAR35	France	Europe	7	RF-1	UNK	UNK
SRA Database	ERR6415201	PICSAR36	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415202	PICSAR38	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415239	PICSAR39	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415203	PICSAR4	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415238	PICSAR42	France	Europe	UNK	UNK	10	RF-1
SRA Database	ERR6415204	PICSAR43	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415205	PICSAR46	France	Europe	9	RF-3	11	RF-2
SRA Database	ERR6415206	PICSAR47	France	Europe	9	RF-3	11	RF-2
SRA Database	ERR6415207	PICSAR48	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415208	PICSAR49	France	Europe	UNK	UNK	10	RF-1
SRA Database	ERR6415209	PICSAR5	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415210	PICSAR51	France	Europe	10	RF-1	12	RF-3
SRA Database	ERR6415211	PICSAR52	France	Europe	10	RF-1	11	RF-2
SRA Database	ERR6415212	PICSAR53	France	Europe	10	RF-1	10	RF-1
SRA Database	ERR6415213	PICSAR55	France	Europe	9	RF-3	10	RF-1
SRA Database	ERR6415214	PICSAR56	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415215	PICSAR58	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415280	PICSAR59	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415216	PICSAR6	France	Europe	7	RF-1	10	RF-1

SRA Database	ERR6415217	PICSAR61	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415218	PICSAR63	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415219	PICSAR64	France	Europe	7	RF-1	UNK	UNK
SRA Database	ERR6415220	PICSAR65	France	Europe	8	RF-2	11	RF-2
SRA Database	ERR6415283	PICSAR67	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415221	PICSAR68	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415222	PICSAR7	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415223	PICSAR70	France	Europe	13	RF-1	11	RF-2
SRA Database	ERR6415224	PICSAR71	France	Europe	UNK	UNK	10	RF-1
SRA Database	ERR6415225	PICSAR72	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415226	PICSAR73	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415227	PICSAR74	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415228	PICSAR75	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415229	PICSAR76	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415230	PICSAR77	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415231	PICSAR78	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415232	PICSAR8	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415233	PICSAR81	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415234	PICSAR85	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415243	PICSAR87	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415235	PICSAR9	France	Europe	9	RF-3	11	RF-2
SRA Database	ERR6415236	PICSAR94	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415241	PICSAR97	France	Europe	9	RF-3	12	RF-3
SRA Database	ERR6415237	PICSAR98	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR4388993	SAMEA7098499	Italy	Europe	7	RF-1	17	RF-2
SRA Database	ERR6415250	SAMEA8989788	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415251	SAMEA8989789	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415252	SAMEA8989790	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415254	SAMEA8989791	France	Europe	11	RF-2	10	RF-1

SRA Database	ERR6415257	SAMEA8989792	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415258	SAMEA8989793	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415259	SAMEA8989794	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415260	SAMEA8989795	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415261	SAMEA8989796	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415262	SAMEA8989797	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415267	SAMEA8989798	France	Europe	10	RF-1	13	RF-1
SRA Database	ERR6415268	SAMEA8989799	France	Europe	10	RF-1	11	RF-2
SRA Database	ERR6415269	SAMEA8989800	France	Europe	7	RF-1	10	RF-1
SRA Database	ERR6415270	SAMEA8989801	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415271	SAMEA8989802	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415272	SAMEA8989803	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415277	SAMEA8989804	France	Europe	10	RF-1	10	RF-1
SRA Database	ERR6415278	SAMEA8989805	France	Europe	7	RF-1	13	RF-1
SRA Database	ERR6415279	SAMEA8989806	France	Europe	9	RF-3	11	RF-2
SRA Database	ERR6415281	SAMEA8989807	France	Europe	7	RF-1	9	RF-3
SRA Database	ERR6415282	SAMEA8989808	France	Europe	7	RF-1	11	RF-2
SRA Database	ERR6415284	SAMEA8989809	France	Europe	12	RF-3	12	RF-3
SRA Database	ERR6415285	SAMEA8989810	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415286	SAMEA8989811	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415287	SAMEA8989812	France	Europe	7	RF-1	12	RF-3
SRA Database	ERR6415295	SAMEA8989813	France	Europe	7	RF-1	10	RF-1
SRA Database	SRR8177401	Telford1	Australia	Oceania	10	RF-1	13	RF-1

[a]Data retrieved either based on WGS of QC/ON samples our group has directly processed, sent for sequencing, and uploaded to the SRA database, sequencing reads from other groups already present within the SRA database, or complete sequences/incomplete assemblies present in the NCBI Nucleotide Database. Highlighted cases had ambiguous repeat sizes within their available assemblies, so manual assembly with their respective SRA records was performed.
[b]Sequence Identifiers based on the Sequence Read Archive ID, or the Genome Assembly ID.

[c]Reading Frames (RF) categorized by repeat size for SSR1. SSR sizes are called as RF-1 (red), RF-2 (yellow), or RF-3 (blue). Instances where SSR size could not be determined are listed as UNK (grey).
[d]Reading Frames (RF) categorized by repeat size for SSR2. SSR sizes are called as RF-1 (green), RF-2 (orange), or RF-3 (purple). Instances where SSR size could not be determined are listed as UNK (grey).

Appendix Table A2: Details of studies on MSIs identified in the current review (n=121)

Title of Publication	Location of Study (Primary)[a]	Total No. of Cases Analyzed (Specimens)	No. (%) of Mixed Infections	Host Species with MSI[b]	Bacterial Species with MSI[c]	Method of Mixed Infection Detection[d]	Specimen or Culture	Total (MSI) HIV Status of Patients	Year of Publication[e]	First Author
Phage Types of *Mycobacterium tuberculosis* in Cultures Isolated from Eskimo Patients	Canada	233 (699)	33 (14.16 %)	*H. sapiens*	*M. tuberculosis*	Phage Typing	Culture	N/A	1975	Mankiewicz
Phage Type of Tubercle Bacilli Isolated from Patients with Two or More Sites of Organ lnvolvement	USA	87 (199)	3 (3.5 %)	*H. sapiens*	*M. tuberculosis*	Phage Typing	Culture	N/A	1976	Bates
Genetic diversity among strains of *Mycobacterium avium* causing monoclonal and polyclonal bacteremia in patients with AIDS	USA	14	2 (14.3 %)	*H. sapiens*	**M. avium* Complex	PFGE	Culture	Positive	1993	Arbeit
Polyclonal infections due to *Mycobacterium avium* complex in patients with AIDS detected by pulsed-field gel electrophoresis of sequential clinical isolates	USA	12 (55)	4 (33.3 %)	*H. sapiens*	**M. avium* Complex	PFGE	Culture	Positive	1994	Slutsky

Mixed-strain infection with a drug-sensitive and multidrug-resistant strain of *Mycobacterium tuberculosis*	Nepal	1	1 (100%)	*H. sapiens*	*M. tuberculosis*	Mixed-linker PCR	Culture	N/A	1995	Theisen
Polyclonal *Mycobacterium avium* Infections in Patients with AIDS: Variations in Antimicrobial Susceptibilities of Different Strains of *M. avium* Isolated from the Same Patient	USA	5 (75)	5 (100%)	*H. sapiens*	*M. avium* Complex	PFGE	Culture	Positive	1995	Von Reyn
Use of different molecular typing techniques for bacteriological follow-up in a clinical trial with AIDS patients with *Mycobacterium avium* bacteremia	France	93	3 (3.2 %)	*H. sapiens*	*M. avium	IS*1245*-RFLP, Strain Specific PCR,*PFGE	Culture	Positive	1997	Picardeau
Computer-assisted analysis of *Mycobacterium avium* fingerprints using insertion elements IS*1245* and IS*1311* in a Caribbean setting	French Guiana, Martinique, Guadeloupe	25 (33)	1 (4.0 %)	*H. sapiens*	*M. avium	IS*1245* Fingerprinting, IS*1311* Fingerprinting	Culture	22 (1) Positive	1997	Devallois

Polyclonal *Mycobacterium avium* complex infections in patients with nodular bronchiectasis	USA	26 (17 Nodular Bronchiectasis Cases, 9 Cavitary Cases)	16 (61.5 %) [15 Nodular Bronchiectasis Cases (88.2 %), 1 Cavitary Case (11.1 %)]	*H. sapiens*	*M. intracellulare* and **M. avium*	PFGE	Culture	Positive	1998	Wallace Jr.
Evidence of exogenous reinfection and mixed infection with more than one strain of *Mycobacterium tuberculosis* among Spanish HIV-infected inmates	Spain	28	1 (3.6 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, pTBN12 Digestion	Culture	28 (1) Positive	1999	Chaves
Simultaneous infection with two drug-susceptible *Mycobacterium tuberculosis* strains in an immunocompetent host	Austria	1	1 (100 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, Spoligotyping	Culture	N/A	1999	Pavlic
Simultaneous infection with two strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis	USA	1	1 (100 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	N/A	1999	Yeh

Genetic diversity of *Mycobacterium avium* recovered from AIDS patients in the Caribbean as studied by a consensus IS*1245*-RFLP method and pulsed-field gel electrophoresis	French Guiana, Martinique, Guadeloupe	31 (45)	2 (6.45 %)	*H. sapiens*	**M. avium*	IS*1245*-RFLP, PFGE	Culture	31 Positive	2000	Legrand
Genotypic patterns of multiple isolates of *M. tuberculosis* from tuberculous HIV patients	Brazil	32	3 (9.38 %)	*H. sapiens*	*M. tuberculosis*	DRE-PCR	Culture	32 (3) Positive	2000	Lourenço
Multiple Isolates from Aids Patients: Aspects of an Analysis by a Genotypic Marker and Antimicrobial Susceptibilities Variations	Brazil	10	2 (20.0 %)	*H. sapiens*	**M. avium*	IS*1245*-RFLP	N/A	10 (2) Positive	2000	Saad

PCR-Restriction Enzyme Analysis of a Bone Marrow Isolate from a Human Immunodeficiency Virus-Positive Patient Discloses Polyclonal Infection with Two *Mycobacterium avium* Strains	Brazil	1	1 (100 %)	*H. sapiens*	**M. avium* Complex	*hsp*65 PCR-restriction enzyme analysis, IS*1245*-RFLP, IS*1311*-RFLP	Culture	Positive	2000	Oliveira
A Molecular Epidemiological Study of *Mycobacterium simiae* Isolated from AIDS Patients in Guadeloupe	Guadeloupe	9	3 (33.3 %)	*H. sapiens*	*M. simiae*	PFGE, *RAPD Analysis	Culture	7 (2) Positive	2000	Legrand
Simultaneous infection with multiple strains of *Mycobacterium tuberculosis*	USA	2	2 (100 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	Negative	2001	Braden

Demonstration of reinfection and reactivation in HIV-negative autopsied cases of secondary tuberculosis: multilesional genotyping of Mycobacterium tuberculosis utilizing IS *6110* and other repetitive element-based DNA fingerprinting	South Africa	13	2 (15.4 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, DR-RFLP, MTB484-RFLP	Culture	Negative	2001	Du Plessis
Multiple *Mycobacterium tuberculosis* Strains in Early Cultures from Patients in a High-Incidence Community Setting	South Africa	131	3 (2.3 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	Negative	2002	Richardson
IS*1311* and IS*1245* Restriction Fragment Length Polymorphism Analyses, Serotypes, and Drug Susceptibilities of *Mycobacterium avium* Complex Isolates Obtained from a Human Immunodeficiency Virus-Negative Patient	Czech Republic	1 (6 sputum isolates)	1 (100%)	*H. sapiens*	**M. avium* Complex	IS*1245*-RFLP, *IS*1311*-RFLP	Culture	Negative	2002	Dvorska

IS*1245* restriction fragment length polymorphism typing of *Mycobacterium avium* from patients admitted to a reference hospital in Campinas, Brazil	Brazil	39 (43)	1 (2.6 %)	*H. sapiens*	**M. avium*	IS*1245*-RFLP	Culture	35 Positive	2003	Panunto
Polyclonal and Compartmentalized Infection by *Mycobacterium tuberculosis* in Patients with Both Respiratory and Extrarespiratory Involvement	Spain	50 (107)	3 (6.0 %)	*H. sapiens*	*M. tuberculosis*	Spoligotyping, DRE-PCR, IS*6110*-RFLP	Culture	42 (3) Positive	2003	García De Viedma
Utility of fast mycobacterial interspersed repetitive unit-variable number tandem repeat genotyping in clinical mycobacteriological analysis	Belgium	3	1 (33.3 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, 12-locus MIRU-VNTR	Culture	N/A	2004	Allix
Simultaneous infection with multiple strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis	India	543	2 (0.4 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	N/A	2004	Das

Genotypic and phenotypic heterogeneity among *Mycobacterium tuberculosis* isolates from pulmonary tuberculosis patients	Bangladesh	97	2 (2.1 %)	*H. sapiens*	*M. tuberculosis*	13-locus MIRU-VNTR, IS*6110*-RFLP	Culture	N/A	2004	Shamputa
Patients with active tuberculosis often have different strains in the same sputum specimen.	South Africa	186	35 (18.8 %)	*H. sapiens*	*M. tuberculosis*	Lineage Specific PCRs, *Spoligotyping	Culture	N/A	2004	Warren
Differential virulence of *Mycobacterium avium* strains isolated from HIV-infected patients with disseminated *M. avium* complex disease	USA	8	2 (25.0 %)	*H. sapiens*	**M. avium* Complex	PFGE	Culture	Positive	2004	Ohkusu
Multiple infection with resistant and sensitive *M. tuberculosis* strains during treatment of pulmonary tuberculosis patients	Peru	10	3 (33.3 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, DRE-PCR	Culture	(1) Positive, 2 (Negative)	2005	Baldeviano-Vidalón
The Beijing genotype and drug resistant tuberculosis in the Aral Sea region of Central Asia	Uzbekistan and Turkmenistan	397	15 (3.8 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, Spoligotyping	Culture	N/A	2005	Cox

Mycobacterium avium restriction fragment lenght polymorphism-IS IS*1245* and the simple double repetitive element polymerase chain reaction typing method to screen genetic diversity in Brazilian strains	Brazil	14 (41) *H. sapiens*, (10) Environmental	2 (14.3 %) *H. sapiens* Only	*H. sapiens, G. gallus domesticus, S. scrofa domesticus, Unknown Vegetable*	**M. avium*	IS*6110*-RFLP, MaDRE-PCR	Culture	14 (2) Positive	2005	De Sequeira
Characterization of clonal complexity in tuberculosis by mycobacterial interspersed repetitive unit–variable-number tandem repeat typing.	Spain	12	1 (8.3 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP and Spoligotyping	Culture	1 Positive	2005	García de Viedma
Reinfection and mixed infection cause changing *Mycobacterium tuberculosis* drug-resistance patterns	South Africa	48	5 (10.4 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP, Strain-Specific PCR	Culture	N/A	2005	Van Rie
Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia	Georgia	199 (597)	26 (13.1 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	N/A	2006	Shamputa

Risk of Acquired Drug Resistance during Short-Course Directly Observed Treatment of Tuberculosis in an Area with High Levels of Drug Resistance	Uzbekistan and Turkmenistan	397	15 (3.8 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	N/A	2007	Cox
Avian tuberculosis in naturally infected captive water birds of the Ardeideae and Threskiornithidae families studied by serotyping, IS*901* RFLP typing, and virulence for poultry	Czech Republic	20 Birds (92 randomly selected *M. avium* subsp. *avium* isolates; 71 tissues and 21 faecal isolates)	1 (5.0 %)	*E. garzetta, B. ibis, E. alba, B. stellaris, T. athiopicus, P. leucorodia*	*M. avium subsp. avium*	IS*901*-RFLP, *IS*1245*-RFLP	Culture	N/A	2007	Dvorska
Molecular investigation of recurrent tuberculosis in patients from Rwanda	Rwanda	710	1/22 (4.5%)	*H. sapiens*	*M. tuberculosis*	Spoligotyping, 12-locus MIRU-VNTR	Culture	N/A	2007	Umubyeyi
Multiple *Mycobacterium tuberculosis* infections in an HIV-infected patient	Netherlands	1	1 (100 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	1 Positive	2007	Van der Zanden

Exogenous Reinfection as a Cause of Multidrug Resistant and Extensively Drug-Resistant Tuberculosis in Rural South Africa	South Africa	17	2 (11.8 %)	*H. sapiens*	*M. tuberculosis*	Spoligotyping	Culture	15 Positive	2008	Andrews
Mixed infections of *Mycobacterium tuberculosis* in tuberculosis patients in Shanghai, China	China	22 (660) VNTR Validation, 249 Mixed Infection Identification	2 (9.1 %), 14 (5.6 %)	*H. sapiens*	*M. tuberculosis*	7-locus VNTR	Culture	N/A	2008	Fang
Mycobacterium avium subsp. *avium* distribution studied in a naturally infected hen flock and in the environment by culture, serotyping and IS*901* RFLP methods	Czech Republic	21 (16 tissue samples tested with RFLP) *G. gallus domesticus*, 35 Environmental Samples	7/16 (43.8 %) *G. gallus domesticus*	*G. gallus domesticus*	*M. avium subsp. avium*	IS*901*-RFLP	Culture	N/A	2008	Shitaye
IS*1245* RFLP-based genotyping study of *Mycobacterium avium* subsp. *hominissuis* isolates from pigs and humans	Slovenia	26 (57) *H. sapiens*, 33 (57) *Sus* Samples	1 (3.8 %) *H. sapiens*, 5 (13.5 %) *Sus*	*H. sapiens, Sus. Spp.*	*M. avium* subsp. *hominissuis*	IS*1245*-RFLP	Culture	1 Positive	2008	Pate

Phenotypic and genotypic variant of MDR-*Mycobacterium tuberculosis* multiple isolates in the same tuberculosis episode, Rio de Janeiro, Brazil	Brazil	1	1 (100 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	1 (1) Positive	2009	Andrade
Mechanisms of heteroresistance to isoniazid and rifampin of *Mycobacterium tuberculosis* in Tashkent, Uzbekistan	Uzbekistan	17	5 (18.5 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR, *IS*6110*-RFLP,*Spoligotyping	Culture	15 Positive	2009	Hofmann-Theil
Penitentiary population of *Mycobacterium tuberculosis* in Kyrgyzstan: Exceptionally high prevalence of the Beijing genotype and its Russia-specific subtype	Kyrgyzstan	56	8 (14.2 %)	*H. sapiens*	*M. tuberculosis*	IS*6110* Inverse PCR , 12-locus MIRU-VNTR, Spoligotyping	N/A	Negative	2009	Mokrousov
High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates	South Africa	(252), (54) Analyzed with MIRU-VNTR	10 isolates (18.5 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR, *Spoligotyping, *IS*6110*-RFLP	Culture	112 isolates Positive	2009	Stavrum

Detection of multiple strains of *Mycobacterium tuberculosis* using MIRU-VNTR in patients with pulmonary tuberculosis in Kampala, Uganda	Uganda	113	8 (7.1 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR	Culture	(3) Positive	2010	Dickman
Mixed Infection with Beijing and Non-Beijing Strains and Drug Resistance Pattern of *Mycobacterium tuberculosis*	Taiwan	185	21 (11.4 %)	*H. sapiens*	*M. tuberculosis*	Lineage Specific PCRs	Culture	1 Positive	2010	Huang
Molecular Detection of Mixed Infections of *Mycobacterium tuberculosis* Strains in Sputum Samples from Patients in Karonga District, Malawi	Malawi	72 patients (160 sputum samples + 377 clinical isolates)	4 (3.2%) sputua. 4 (1.1%) isolates (Spoligotype only)	*H. sapiens*	*M. tuberculosis*	Strain Specific PCR, *24-locus MIRU-VNTR	Specimen and Culture	37/59 Positive	2010	Mallard
Diversity of *Mycobacterium tuberculosis* genotypes circulating in Ndola, Zambia	Zambia	156	5 (3.2 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR	Culture	N/A	2010	Mulenga

Mixed-strain *Mycobacterium tuberculosis* infections among patients dying in a hospital in KwaZulu-Natal, South Africa	South Africa	56	5 (8.9 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	54 (5) Positive	2011	Cohen
Whole-Genome Sequencing and Social Network Analysis of a Tuberculosis Outbreak	Canada	32	4 (12.5 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis, 24-locus MIRU-VNTR, IS*6110*-RFLP	Culture	Negative	2011	Gardy
Systematic survey of clonal complexity in tuberculosis at a populational level and detailed characterization of the isolates involved	Spain	703 Respiratory + 71 Respiratory extraoulmonary = 774	9 (1.2 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR, 24-locus MIRU-VNTR, IS*6110*-RFLP	Culture	N/A	2011	Navarro
Mixed infection with Beijing and non-Beijing strains in pulmonary tuberculosis in Taiwan: prevalence, risk factors, and dominant strain	Taiwan	466	14 (3.0 %)	*H. sapiens*	*M. tuberculosis*	Lineage Specific PCRs	Culture	N/A	2011	Wang
Disseminated tuberculosis in human immunodeficiency virus infection: ineffective immunity, polyclonal disease and high mortality	Tanzania	1955 Non-disseminated TB, 20 Disseminated TB (Typed with IS*6110*-RFLP)	3/20 (15.0 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	Positive	2011	Von Reyn

Molecular epidemiology of HIV-associated tuberculosis in Dar es Salaam, Tanzania: strain predominance, clustering, and polyclonal disease	Tanzania	14	6 (42.9 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	14 (6) Positive	2012	Adams
A first insight on the population structure of *Mycobacterium tuberculosis* complex as studied by spoligotyping and MIRU-VNTRs in Bogota, Colombia	Colombia	146	1 (0.7 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR, IS*6110*-RFLP	Culture	1 (1) Positive	2012	Cerezo
DNA Typing of *Mycobacterium bovis* Isolates from Badgers (*Meles meles*) Culled from Areas in Ireland with Different Levels of Tuberculosis Prevalence	Republic of Ireland	52 (93)	3 (5.8 %)	*M. meles*	*M. bovis*	6-locus VNTR, Spoligotyping	Culture	N/A	2012	Furphy
Mixed tuberculosis infections in rural South Vietnam	Vietnam	1248	39 (3.1 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR, Spoligotyping, IS*6110*-RFLP	Culture	N/A	2012	Huyen

Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study	Malaysia, South Africa, Thailand	47	6 (12.8 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis	Culture	Negative	2013	Bryant
Metagenomic Analysis of Tuberculosis in a Mummy	Hungary	1	1 (100%)	*H. sapiens*	*M. tuberculosis*	WGS (Metagenomics)	Specimen	N/A	2013	Chan
Population structure of mixed *Mycobacterium tuberculosis* infection is strain genotype and culture medium dependent	South Africa	206	31 (15.1 %)	*H. sapiens*	*M. tuberculosis*	Strain Specific PCR	Culture	N/A	2013	Hanekom
Undetected Multidrug-Resistant Tuberculosis Amplified by First-line Therapy in Mixed Infection	Portugal	1	1 (100 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	Negative	2013	Hingley-Wilson
Molecular investigation of multiple strain infections in patients with tuberculosis in Mubende district, Uganda	Uganda	72	8 (11.1 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR	Culture	48/74 (65%) positive	2013	Muwonge

Multiple samples improve the sensitivity for detection of mixed *Mycobacterium* infections	China	89	10 (11.2 %)	*H. sapiens*	*M. tuberculosis*	Strain Specific PCR, 16-locus VNTR	Culture	N/A	2013	Peng
Association between polyclonal and mixed mycobacterial *Mycobacterium avium* complex infection and environmental exposure	Japan	120 (94 *M. avium* subsp. *avium*, 26 *M. intracellulare*)	27 (22.5 %), 20 *M. avium* subsp. *avium* (21.3 %), 7 *M. intracellulare* (26.9 %)	*H. sapiens*	*M. avium subsp. avium and M. intracellulare*	15-locus MIRU-VNTR (*M. avium*), 16-locus MIRU-VNTR (*M. intracellulare*)	Culture	Negative	2014	Fujita
Multi-locus variable-number tandem repeat analysis (MLVA) reveals heterogeneity of *Mycobacterium bovis* strains and multiple genotype infections of cattle in Ethiopia	Ethiopia	26	10 (38.5 %)	*B. taurus*	*M. bovis*	MLVA 28-locus VNTR, Spoligotyping	Culture	N/A	2014	Biffa
Contribution of spoligotyping and MIRU-VNTRs to characterize prevalent *Mycobacterium tuberculosis* genotypes infecting tuberculosis patients in Morocco	Morocco	208 (219)	4 (1.9 %)	*H. sapiens*	*M. tuberculosis*	12-locus MIRU-VNTR, Spoligotyping	Culture	N/A	2014	Chaoui

Multiple strain infections and high genotypic diversity among *Mycobacterium avium* subsp. *paratuberculosis* field isolates from diseased wild and domestic ruminant species in the eastern Alpine region of Austria	Austria	37	2 (5.4 %)	*C. capreoulus, C. elaphus, R. rupicapra, O. orientalis musimon, C. ibex*	*M. avium subsp. paratuberculosis*	8-locus MIRU-VNTR	Culture	N/A	2014	Gerritsmann
Genotypic characterization by spoligotyping and VNTR typing of *Mycobacterium bovis* and *Mycobacterium caprae* isolates from cattle of Tunisia	Tunisia	48 (35)	1 (2.08 %)	*B. taurus*	*M. bovis*	6-locus VNTR, Spoligotyping	Culture	N/A	2014	Lamine-Khemiri
Genotyping of ancient *Mycobacterium tuberculosis* strains reveals historic genetic diversity	Britain/France	10	1 (10.0 %)	*H. sapiens*	*M. tuberculosis*	*gyr*A-PCR	Specimen	N/A	2014	Müller

Mixed *Mycobacterium tuberculosis* Complex Infections and False-Negative Results for Rifampin Resistance by GeneXpert MTB/RIF Are Associated with Poor Clinical Outcomes	USA	370	37 (10.0 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	N/A	2014	Zetola
Mycobacterium avium subsp. *hominissuis* Infection in Swine Associated with Peat Used for Bedding	Norway	46 *Sus* Samples, 23 Environmental Samples, 16 Peat Bedding Samples	8/20 (40.0 %) *Sus* Samples with isolates from more than one organ	*Sus* spp.	*M. avium* subsp. *hominissuis*	8-locus MATR-VNTR	Culture	N/A	2014	Johansen
Polyclonality among clinical strains of nonpigmented rapidly growing mycobacteria: phenotypic and genotypic differences and their potential implications	Spain	64	13 (20.3 %)	*H. sapiens*	*M. abscessus, M. chelonae, M. fortuitum, M. mucogenicum*	RAPD Analysis	Culture	N/A	2015	García-Pedrazuela

Predominant *Mycobacterium tuberculosis* Families and High Rates of Recent Transmission among New Cases Are Not Associated with Primary Multidrug Resistance in Lima, Peru	Peru	844	12 (1.4 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR	Culture	N/A	2015	Barletta
Molecular typing of Argentinian *Mycobacterium avium* subsp. *paratuberculosis* isolates by multiple-locus variable number-tandem repeat analysis	Argentina	(97) (78 *B. taurus*, 16 *C. aegagrus hircus*, 3 *O. aries*)	1 *B. taurus* (N/A)	*B. taurus, C. aegagrus hircus, O. aries*	*M. avium subsp. paratuberculosis*	8-locus MIRU-VNTR	Specimen	N/A	2015	Gioffré
Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe	Hungary	8	5 (62.50 %)	*H. sapiens*	*M. tuberculosis*	WGS (Metagenomics)	Specimen	N/A	2015	Kay
Discrepancies in Drug Susceptibility Test for Tuberculosis Patients Resulted from the Mixed Infection and the Testing System	China	(20 paired isolates)	3/12 (25 %)	*H. sapiens*	*M. tuberculosis*	12-MIRU-VNTR	Culture	N/A	2015	Mei

Multiple sampling and discriminatory fingerprinting reveals clonally complex and compartmentalized infections by *M. bovis* in cattle	Spain	55	6 (10.0 %)	*B. taurus*	*M. bovis*	9-locus MIRU-VNTR, 24-locus MIRU-VNTR, Spoligotyping	Culture	N/A	2015	Navarro
Prevalence and risk factors of mixed *Mycobacterium tuberculosis* complex infections in China	China	3248	48 (1.5 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	N/A	2015	Pang
Typing of *Mycobacterium avium* subspecies *paratuberculosis* isolates from Newfoundland using fragment analysis	Canada	18 (85)	18 (100 %)	*B. taurus*	*M. avium subsp. paratuberculosis*	4-locus SSR typing (VNTR)	Culture	N/A	2015	Podder
Advanced immune suppression is associated with increased prevalence of mixed-strain *Mycobacterium tuberculosis* infections among persons at high risk for drug-resistant tuberculosis in Botswana	Botswana	370	37 (10.0 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	279 (32) Positive	2015	Shin

High Genotypic Discordance of Concurrent *Mycobacterium tuberculosis* Isolates from Sputum and Blood of HIV-Infected Individuals	Uganda	51	2 (3.9 %), Potentially 26 others (51.0 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR, Spoligotyping	Culture	Positive	2015	Ssengooba
Mycobacterium tuberculosis polyclonal infections and microevolution identified by MIRU-VNTRs in an epidemiological study	Guyana and Suriname	161	1 (0.6 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR, *Spoligotyping	Culture	(1) Positive	2015	Streit
An investigation on the population structure of mixed infections of *Mycobacterium tuberculosis* in Inner Mongolia, China	China (Mongolia)	384	12(3.1 %)	*H. sapiens*	*M. tuberculosis*	Spoligotyping, 15-locus MIRU-VNTR	Culture	N/A	2015	Wang
Mixed Infections and Rifampin Heteroresistance among *Mycobacterium tuberculosis* Clinical Isolates	China	499	12 (2.4 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Specimen	N/A	2015	Zheng

Outbreak of tuberculosis in a closed setting: views on transmission based on results from molecular and conventional methods	Russia	7	1 (14.3 %)	*H. sapiens*	*M. tuberculosis*	15-locus MIRU-VNTR	Culture	Negative	2016	Antusheva
Examination of *Mycobacterium avium* subspecies *paratuberculosis* mixed genotype infections in dairy animals using a whole genome sequencing approach	Canada	4 (6)	1 (25.0 %)	*B. taurus*	*M. avium subsp. paratuberculosis*	WGS Analysis	Culture	N/A	2016	Davidson
Deep Whole-Genome Sequencing to Detect Mixed Infection of *Mycobacterium tuberculosis*	Various	782	47 (6.0 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis	Culture	N/A	2016	Gan
High prevalence of *Mycobacterium tuberculosis* mixed infection in the capital of moderate tuberculosis incidence country	Iran	75	20 (26.7 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	Negative	2016	Hajimiri

Drug resistance characteristics and cluster analysis of *M. tuberculosis* in Chinese patients with multiple episodes of anti-tuberculosis treatment	China	166	32 (19.3 %)	*H. sapiens*	*M. tuberculosis*	28-locus MIRU-VNTR	Culture	N/A	2016	Hu
Emergence of mixed infection of Beijing/Non-Beijing strains among multi-drug resistant *Mycobacterium tuberculosis* in Pakistan	Pakistan	102	15 (14.7 %)	*H. sapiens*	*M. tuberculosis*	Lineage Specific PCRs	Culture	N/A	2016	Mustafa
Outcomes, infectiousness, and transmission dynamics of patients with extensively drug-resistant tuberculosis and home-discharged patients with programmatically incurable tuberculosis: a prospective cohort study	South Africa	273	11 (4.0 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis and *IS*6110*-RFLP	Culture	N/A	2017	Dheda
Molecular epidemiology of *Mycobacterium bovis* in Cameroon	Cameroon	207 (317)	32 (15.5 %)	*B. taurus*	*M. bovis*	24-locus MIRU-VNTR, Spoligotyping	Culture	N/A	2017	Egbe

Bias in detection of *Mycobacterium tuberculosis* polyclonal infection: Use clinical samples or cultures?	Iran	38	21 (55.3 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	N/A	2017	Farmanfarmaei
Tuberculosis in Swiss captive Asian elephants: microevolution of *Mycobacterium tuberculosis* characterized by multilocus variable-number tandem-repeat analysis and whole-genome sequencing	Switzerland	3	1 (33.3 %)	*E. maximus*	*M. tuberculosis*	26-locus MIRU-VNTR, *Spoligotyping, *WGS Analysis	Culture	N/A	2017	Ghielmetti
Evaluation of the impact of polyclonal infection and heteroresistance on treatment of tuberculosis patients	Iran	66 (98 clinical samples, 288 colonies)	35 (53.03 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Specimen	N/A	2017	Kamakoli
Genetic diversity of multidrug-resistant *Mycobacterium tuberculosis* strains isolated from tuberculosis patients in Iran using MIRU-VNTR technique	Iran	88 (88)	1 (1.13 %)	*H. sapiens*	*M. tuberculosis*	12-locus MIRU-VNTR	Culture	N/A	2017	Khosravi

Tuberculosis cases caused by heterogeneous infection in Eastern Europe and their influence on outcomes	Lithuania, Latvia, Russia	220 (512)	13 (5.91 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	13 Positive	2017	Kontsevaya
Polyclonal Pulmonary Tuberculosis Infections and Risk for Multidrug Resistance, Lima, Peru	Peru	3098	161 (5.2 %)	*H. sapiens*	*M. tuberculosis*	24-MIRU-VNTR	Culture	108 (3.5%) Positive	2017	Nathavitharana
Whole-genome analysis of mycobacteria from birds at the San Diego Zoo	USA	113 (132), 49 birds with *M. avium* species, 48 birds with *M. genavense* species	12 total (10.6 %), 2 *M. avium* (4.1 %), 10 *M. genavense* (20.8 %)	*Multiple Unknown Bird Species*	*M. avium subsp. avium and M. genavense*	WGS Analysis	Culture	N/A	2017	Pfeiffer
Pros and cons of direct genotyping on tuberculosis clinical samples	Iran	32	25 (79.1 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	N/A	2017	Sadegh
Use of whole-genome sequencing to distinguish relapse from reinfection in a completed tuberculosis clinical trial	South Africa, Zambia, Botswana and Zimbabwe	(36 paired isolates)	1 (2.8 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis, *24-locus MIRU-VNTR, **in-silico* Spoligotyping	Culture	N/A	2017	Witney

Mixed *Mycobacterium tuberculosis*-Strain Infections Are Associated With Poor Treatment Outcomes Among Patients With Newly Diagnosed Tuberculosis, Independent of Pretreatment Heteroresistance	Botswana	260	21 (8.1 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	N/A	2018	Shin
Challenge in direct Spoligotyping of *Mycobacterium tuberculosis*: a problematic issue in the region with high prevalence of polyclonal infections	Iran	14	8 (57.1 %)	*H. sapiens*	*M. tuberculosis*	Spoligotyping	Culture	N/A	2018	Kamakoli
Drug-Susceptible and Multidrug-Resistant *Mycobacterium tuberculosis* in a Single Patient	USA	1	1 (100 %)	*H. sapiens*	*M. tuberculosis*	Spoligotyping, WGS Analysis	Culture	1 (1) Positive	2019	Baffoe-Bonnie

Usefulness of 3'- 5' IS*6110*-RFLP genotyping and spoligotyping of *Mycobacterium tuberculosis* isolated in a tertiary hospital: a retrospective study detecting unsuspected epidemiological events	Brazil	106	3 (2.8 %)	*H. sapiens*	*M. tuberculosis*	IS*6110*-RFLP	Culture	N/A	2019	De Almeida
Whole genome sequencing provides additional insights into recurrent tuberculosis classified as endogenous reactivation by IS6110 DNA fingerprinting	South Africa	25	2 (8.0 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis	Culture	1 Positive, 11 Negative, 13 Unknown	2019	Dippenaar
Comparison of MIRU-VNTR genotyping between old and fresh clinical samples in tuberculosis	Iran	14 old samples, 9 new samples	0 (0.00 %), 4 (44.44 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Specimen and Culture	N/A	2019	Kamakoli
Early detection of emergent extensively drug-resistant tuberculosis by flow cytometry-based phenotyping and whole-genome sequencing	South Africa	20	1/20 (5.00 %)	*H. sapiens*	*M. tuberculosis*	*φ2GFP10 phage Flow Cytometry, WGS Analysis	Culture	N/A	2019	O'Donnell

Genome sequencing of *Mycobacterium pinnipedii* strains: Genetic characterization and evidence of superinfection in a South American sea lion (*Otaria flavescens*)	Peru	1	1 (100 %)	*O. flavescens*	*M. pinnipedii*	Spoligotyping, 24-locus MIRU-VNTR, WGS Analysis	Culture	N/A	2019	Silva-Pereira
Diversity of *Mycobacterium tuberculosis* in the Middle Fly District of Western Province, Papua New Guinea: microbead-based spoligotyping using DNA from Ziehl-Neelsen-stained microscopy preparations	Papua New Guinea	162, 80 typed with Spoligotyping	20/80 (25.0 %)	*H. sapiens*	*M. tuberculosis*	Spoligotyping	Specimen	N/A	2019	Guernier-Cambert
Retrospective evaluation of natural course in mild cases of *Mycobacterium avium* complex pulmonary disease	Japan	65 (50 Observation, 15 Treatment)	9 (13.8 %), [4 Observation (8.0 %), 5 Treatment (33.3 %)]	*H. sapiens*	**M. avium* Complex	16-locus MATR-VNTR (*M.avium*), 5-locus HNTR (VNTR) (*M. avium*), *16-locus VNTR (*M. intracellulare*)	Culture	N/A	2019	Kimizuka

Screening of inmates transferred to Spain reveals a Peruvian prison as a reservoir of persistent *Mycobacterium tuberculosis* MDR strains and mixed infections	Spain	6	1 (16.7 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	N/A	2020	Abascal
QuantTB - a method to classify mixed *Mycobacterium tuberculosis* infections within whole genome sequencing data	Malaysia, South Africa, Thailand	50	4 (8.0 %)	*H. sapiens*	*M. tuberculosis*	QuanTB WGS Analysis	Culture	Negative	2020	Anyansi
Possible Transmission Mechanisms of Mixed *Mycobacterium tuberculosis* Infection in High HIV Prevalence Country, Botswana	Botswana	2051	34 (1.4 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	(18) Positive, (14) Negative, (2) Unknown	2020	Baik
Prediction of the hidden genotype of mixed infection strains in Iranian tuberculosis patients	Iran	45	25 (55.6 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Culture	Negative	2020	Kamakoli

Application of MIRU–VNTR on smear slides: a shortcut for detection of polyclonal infections in tuberculosis patients	Iran	14 (14 Smear Slides, 14 Clinical Specimens, 14 Cultures)	6 Smear Slides (42.86 %), 8 Clinical Specimens (57.14 %), 1 Culture (7.14 %)	*H. sapiens*	*M. tuberculosis*	24-locus MIRU-VNTR	Specimen and Culture	N/A	2020	Kamakoli
Dynamics of within-host *Mycobacterium tuberculosis* diversity and heteroresistance during treatment	South Africa	352	21 (6.00 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis	Culture	N/A	2020	Nimmo
Tracing cross species transmission of *Mycobacterium bovis* at the wildlife/livestock interface in South Africa	South Africa	(51) *B. taurus*, (41) *S. caffer*, (4) *P. leo*, (2) *Papio spp.*, (1) *P. africanus*	1 *B. taurus* MGI (UNK %)	*S. caffer, B. taurus, P. leo, Papio spp, P. africanus*	*M. bovis*	13-locus MIRU-VNTR	Culture	N/A	2020	Sichewo
A retrospective genomic analysis of drug-resistant strains of *M. tuberculosis* in a high-burden setting, with an emphasis on comparative diagnostics and reactivation and reinfection status	Moldova	278	4 (1.4 %)	*H. sapiens*	*M. tuberculosis*	WGS Analysis	Culture	N/A	2020	Wollenberg

[a]*Study locations were based on the origin of the samples. Study locations are coloured by continent, grouped as African (orange), Asian (purple), Europe (light blue), North American (red), South American (light green), and Oceanian (dark blue). Studies with samples from more than one continent were coloured dark green.*

[b]*Host identities are coloured based on whether samples originated from humans (yellow), non-human animals (purple), or both (grey).*
[c]*Discrimination of MAC or M. avium subspecies not performed, possibility of mixed species infection exists. Bacterial species are coloured based on whether they are classified as members of the MTBC (orange) or NTM (blue).*
[d]*Methods marked with an "*" were unable to identify MSIs.*
[e]*Year each study was conducted is coloured based on whether it was conducted Pre-2000 (red), from 2000-2009 (green), from 2010-2019 (blue), or in 2020 (purple).*

Appendix Table A3: Assembly statistics for MAP isolates (n=67) from Quebec and Ontario Herds

Strain ID	Herd ID	Province	Genome Size (bp)	GC Content (%)	Average Coverage (X)	# Contigs	Average Contig Size	N50
23	QC-1	QC	4781165	69.33%	18	79	32434.6	118495
32	QC-1	QC	4779851	69.33%	13	81	34752.6	118495
117	ON-1	ON	4780632	69.33%	16	83	29830.5	128709
137	ON-1	ON	4779052	69.33%	14	78	33780.2	118498
160	QC-2	QC	4781128	69.33%	12	83	33321.4	118496
171	QC-3	QC	4781233	69.33%	15.5	73	30997.6	131207
183	QC-2	QC	4780754	69.33%	13	78	34271.9	128709
191	QC-2	QC	4781072	69.33%	16	85	31784.2	118164
206	QC-3	QC	4781381	69.33%	16.5	77	32442.6	122818
211	ON-2	ON	4780009	69.33%	17	77	29839.1	122818
219	QC-1	QC	4781551	69.33%	14.5	76	36069.7	131207
233	ON-1	ON	4780255	69.33%	13	86	34009.7	118071
236	ON-1	ON	4780663	69.33%	12.5	89	30377.5	106432
242	ON-1	ON	4780978	69.33%	16	74	27639.3	131207
252	ON-1	ON	4780633	69.33%	13	80	36331.1	118164
325	ON-3	ON	4767492	69.33%	14	81	31096.9	118495
326	ON-3	ON	4786390	69.33%	18.5	76	28482.3	128697
342	ON-4	ON	4782026	69.33%	18	78	30221.6	122818
345	ON-4	ON	4787043	69.33%	16	78	27208	118495
366	QC-1	QC	4780530	69.33%	15	78	32218.4	118495
375	QC-4	QC	4787090	69.34%	17	77	32482	122818
392	QC-5	QC	4786337	69.34%	14	78	33141.7	118495
400	QC-4	QC	4786604	69.33%	15	82	32470.2	118164
418	QC-3	QC	4781018	69.33%	15	81	30787	118496
441	QC-6	QC	4786572	69.34%	17	80	33831.6	122816
464	ON-5	ON	4780427	69.33%	13	77	33558.4	128712

500	QC-7	QC	4781571	69.33%	16	74	33809.1	128709
505	QC-7	QC	4779206	69.33%	16	81	30388.3	118164
508	QC-8	QC	4785715	69.33%	8	84	36357.9	118041
515	QC-9	QC	4787793	69.34%	16	77	34326.6	122816
516	QC-9	QC	4786333	69.33%	16	82	33132.6	118164
531	QC-10	QC	4787797	69.34%	15	91	31407.1	106431
552	ON-2	ON	4781331	69.33%	17.5	73	35813.4	124763
560	ON-2	ON	4787009	69.33%	17	85	29869	118164
586	ON-6	ON	4780734	69.33%	12.5	88	35510	106432
594	ON-6	ON	4786810	69.33%	12.5	88	35296.1	102085
631	ON-7	ON	4786285	69.33%	14	79	35313.5	118165
638	ON-7	ON	4780241	69.33%	12.5	81	33320.3	111104
648	ON-7	ON	4779900	69.33%	16.5	84	31990.5	118469
863	ON-4	ON	4780994	69.33%	15	79	35530.6	122818
869	QC-4	QC	4785898	69.34%	16.5	87	28797.2	111129
878	QC-4	QC	4786648	69.33%	15	81	35056.9	118437
885	QC-5	QC	4786162	69.33%	16	92	30800.2	99340
897	ON-5	ON	4780095	69.33%	17	83	31784.3	118154
910	ON-8	ON	4787058	69.33%	13	78	35321.9	118493
918	ON-8	ON	4787128	69.33%	15	80	35574.7	118493
928	ON-8	ON	4794312	69.33%	14.5	84	25518.4	118579
938	QC-10	QC	4786667	69.34%	13	81	32688.4	122818
948	QC-10	QC	4787670	69.34%	14	81	35060.7	118164
1007	ON-6	ON	4786380	69.33%	13.5	85	35300	111105
1023	ON-6	ON	4786749	69.33%	12.5	77	34319.8	118433
1032	QC-11	QC	4787318	69.33%	13.5	86	33358.5	118164
1044	QC-11	QC	4786152	69.34%	14	84	36917.2	118164
1071	QC-5	QC	4786936	69.33%	16.5	84	29330.7	118495

1144	QC-7	QC	4781871	69.33%	16	82	35268.9	122816
1157	QC-8	QC	4786071	69.34%	15.5	85	31007.7	114281
1240	ON-3	ON	4787247	69.33%	13	74	35591.2	122813
1402	ON-3	ON	4787104	69.33%	12.5	77	36381.5	128697
1409	ON-5	ON	4780899	69.33%	16	75	34770.8	131207
1419	ON-8	ON	4786295	69.33%	14.5	79	32693.1	118493
1427	QC-9	QC	4787137	69.34%	15.5	75	32930	125592
1438	QC-8	QC	4785625	69.34%	14.5	77	32260.1	122818
1452	ON-2	ON	4786015	69.34%	13.5	80	37795	118493
1495	ON-6	ON	4779823	69.33%	16	80	34019.6	118498
1512	ON-6	ON	4787075	69.33%	15.5	79	34075.9	118433
1517	QC-11	QC	4787035	69.34%	17.5	80	30439.9	122818
1578	QC-12	QC	4785883	69.33%	19	87	30611.9	118164
	Average		4783833.70	69.33%	14.93	80.66	32882.38	119603.09

Appendix Table A4: CheckM Results to Determine Assembly Completeness and Contamination in MAP isolates (n=67) from Quebec and Ontario Herds and select reference genomes from NCBI (n=7)

Strain ID	Marker Lineage	# genomes	# markers	# marker sets	0	1	2	3	4	5+	Completeness	Contamination	Strain Heterogeneity
23	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4	50
32	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.39	50
117	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	1.11	75
137	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.36	0
160	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.42	50
171	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
183	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.67	50
191	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
206	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
211	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.73	75
219	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.88	50
233	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.27	0.33	0
236	g_Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	1.28	80
242	g_Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	0.85	60
252	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
325	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
326	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.5	66.67

342	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5	0
345	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.37	50
366	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5	50
375	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.45	66.67
392	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
400	g_Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	0.93	60
418	g_Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	0.83	80
441	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.64	75
464	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
500	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	39	50
505	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
508	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.83	75
515	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	1	66.67
516	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
531	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
552	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
560	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.56	75
586	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.38	50
594	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
631	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.37	50

638	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.5	66.67
648	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.83	66.67
863	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
869	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.5	66.67
878	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
885	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	1.17	50
897	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.78	66.67
910	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.67	66.67
918	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	1	66.67
928	g_Mycobacterium (UID1816)	100	693	300	4	687	2	0	0	0	99.21	0.4	50
938	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.39	66.67
948	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.67	50
1007	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
1023	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5	50
1032	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
1044	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
1071	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.27	0.56	66.67
1144	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.58	66.67
1157	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
1240	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	1.33	75

1402	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
1409	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
1419	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	1	33.33
1427	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44	50
1438	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
1452	g_Mycobacterium (UID1816)	100	693	300	2	689	2	0	0	0	99.61	0.67	50
1495	g_Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.47	66.67
1512	g_Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5	50
1517	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
1578	g_Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.78	75
K-10	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
NL_89C	g_Mycobacterium (UID1816)	100	693	300	2	690	1	0	0	0	99.61	0.33	0
NL_93B	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
NL_95A	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
NL_95B	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
NL_95E	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0
NL_96E	g_Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33	0

Appendix Table A5: Snippy Results of MAP strains (n=77) using type II strain K-10 as a reference sequence

Strain ID	Province	Herd	SNPs	MNPs	INS	DEL	COMPLEX	TOTAL
23	QC	QC-1	100	3	7	43	3	156
32	QC	QC-1	100	2	8	43	5	158
117	ON	ON-1	86	1	9	41	5	142
137	ON	ON-1	84	1	10	41	5	141
160	QC	QC-2	94	1	9	42	7	153
171	QC	QC-3	91	1	9	44	6	151
183	QC	QC-2	94	2	9	45	5	155
191	QC	QC-2	96	1	10	43	6	156
206	QC	QC-3	98	1	10	42	6	157
211	ON	ON-2	86	2	13	41	5	147
219	QC	QC-1	99	1	8	43	6	157
233	ON	ON-1	71	0	11	41	7	130
236	ON	ON-1	72	1	10	43	6	132
242	ON	ON-1	84	2	9	40	4	139
252	ON	ON-1	70	3	9	42	4	128
325	ON	ON-3	121	2	10	46	5	184
326	ON	ON-3	126	3	9	46	5	189
342	ON	ON-4	104	3	10	42	4	163
345	ON	ON-4	122	2	8	43	6	181
366	QC	QC-1	96	2	7	42	5	152
375	QC	QC-4	127	1	11	47	5	191
392	QC	QC-5	120	0	10	51	11	192
400	QC	QC-4	127	1	12	45	6	191
418	QC	QC-3	93	2	10	42	5	152
441	QC	QC-6	126	2	13	46	6	193

464	ON	ON-5	106	3	11	42	4	166
500	QC	QC-7	99	2	9	42	5	157
505	QC	QC-7	101	2	9	43	5	160
508	QC	QC-8	119	2	9	45	9	184
515	QC	QC-9	124	2	12	46	5	189
516	QC	QC-9	121	2	11	48	5	187
531	QC	QC-10	115	1	10	43	5	174
552	ON	ON-2	84	1	12	40	6	143
560	ON	ON-2	128	2	12	43	7	192
586	ON	ON-6	92	1	13	41	6	153
594	ON	ON-6	126	2	13	46	4	191
631	ON	ON-7	125	3	13	44	5	190
638	ON	ON-7	89	2	9	43	7	150
648	ON	ON-7	90	0	10	44	9	153
863	ON	ON-4	99	2	10	42	4	157
869	QC	QC-4	127	1	11	46	6	191
878	QC	QC-4	127	2	11	45	5	190
885	QC	QC-5	123	1	11	49	10	194
897	ON	ON-5	108	1	11	39	6	165
910	ON	ON-8	132	1	14	44	5	196
918	ON	ON-8	129	1	12	44	6	192
928	ON	ON-8	51	1	6	39	5	102
938	QC	QC-10	114	2	11	43	5	175
948	QC	QC-10	111	1	9	44	5	170
1007	ON	ON-6	76	1	7	42	5	131
1023	ON	ON-6	127	2	12	45	5	191
1032	QC	QC-11	123	0	10	48	11	192

1044	QC	QC-11	125	1	11	48	10	195
1071	QC	QC-5	120	2	10	50	9	191
1144	QC	QC-7	98	2	9	44	5	158
1157	QC	QC-8	115	2	10	48	9	184
1240	ON	ON-3	127	3	9	44	5	188
1402	ON	ON-3	125	2	8	44	6	185
1409	ON	ON-5	104	3	11	40	4	162
1419	ON	ON-8	132	2	13	45	4	196
1427	QC	QC-9	124	3	11	43	4	185
1438	QC	QC-8	117	1	10	45	10	183
1452	ON	ON-2	84	1	8	46	5	144
1495	ON	ON-6	96	2	10	40	5	153
1512	ON	ON-6	130	2	13	46	4	195
1517	QC	QC-11	122	2	10	47	9	190
1578	QC	QC-12	132	1	12	44	8	197
K-10	N/A	NCBI	0	0	0	0	0	0
MAPK_JB15/16	N/A	NCBI	641	0	47	98	13	799
NL 89C	NL	NCBI	219	0	19	56	9	303
NL 93B	NL	NCBI	99	0	9	43	7	158
NL 95A	NL	NCBI	122	0	9	42	7	180
NL 95B	NL	NCBI	100	0	8	44	7	159
NL 95E	NL	NCBI	111	0	13	48	8	180
NL 96E	NL	NCBI	97	0	8	43	7	155
S397	N/A	NCBI	3358	0	161	252	91	3862
Telford	N/A	NCBI	3381	0	167	234	85	3867

Table A6: Snippy Results of MAP strains (n=77) using type I strain Telford as a reference sequence

Strain ID	Province	Herd	SNPs	MNPs	INS	DEL	COMPLEX	TOTAL
23	QC	QC-1	3422	28	169	169	65	3853
32	QC	QC-1	3418	26	167	167	64	3842
117	ON	ON-1	3412	27	169	170	64	3842
137	ON	ON-1	3401	26	163	170	66	3826
160	QC	QC-2	3405	27	165	168	62	3827
171	QC	QC-3	3415	28	165	171	64	3843
183	QC	QC-2	3416	25	163	172	68	3844
191	QC	QC-2	3413	28	168	172	62	3843
206	QC	QC-3	3415	23	173	173	69	3853
211	ON	ON-2	3409	31	171	171	63	3845
219	QC	QC-1	3423	27	166	171	69	3856
233	ON	ON-1	3403	26	168	169	66	3832
236	ON	ON-1	3396	29	161	173	62	3821
242	ON	ON-1	3408	28	171	168	64	3839
252	ON	ON-1	3402	23	166	172	69	3832
325	ON	ON-3	3391	24	159	171	65	3810
326	ON	ON-3	3418	30	162	173	61	3844
342	ON	ON-4	3426	25	165	171	66	3853
345	ON	ON-4	3416	26	166	170	65	3843
366	QC	QC-1	3418	28	170	166	67	3849
375	QC	QC-4	3418	29	166	173	59	3845
392	QC	QC-5	3412	22	168	175	73	3850
400	QC	QC-4	3414	25	171	173	62	3845
418	QC	QC-3	3415	27	170	170	64	3846

441	QC	QC-6	3418	30	172	174	62	3856
464	ON	ON-5	3423	23	168	169	67	3850
500	QC	QC-7	3421	27	170	172	64	3854
505	QC	QC-7	3421	27	167	173	65	3853
508	QC	QC-8	3392	17	162	169	73	3813
515	QC	QC-9	3415	23	169	173	67	3847
516	QC	QC-9	3410	22	171	173	72	3848
531	QC	QC-10	3419	26	165	170	64	3844
552	ON	ON-2	3406	28	169	170	62	3835
560	ON	ON-2	3423	28	172	172	62	3857
586	ON	ON-6	3411	28	167	167	61	3834
594	ON	ON-6	3407	23	167	170	68	3835
631	ON	ON-7	3411	27	168	173	67	3846
638	ON	ON-7	3402	24	160	170	66	3822
648	ON	ON-7	3415	29	164	172	68	3848
863	ON	ON-4	3417	26	168	170	63	3844
869	QC	QC-4	3415	23	171	170	68	3847
878	QC	QC-4	3416	21	168	172	67	3844
885	QC	QC-5	3417	24	170	173	76	3860
897	ON	ON-5	3429	24	170	167	72	3862
910	ON	ON-8	3418	28	165	172	63	3846
918	ON	ON-8	3415	28	172	173	64	3852
928	ON	ON-8	3365	24	164	165	69	3787
938	QC	QC-10	3417	28	164	170	62	3841
948	QC	QC-10	3418	31	161	168	59	3837
1007	ON	ON-6	3361	25	164	166	67	3783
1023	ON	ON-6	3409	25	173	175	66	3848

1032	QC	QC-11	3413	29	165	174	64	3845
1044	QC	QC-11	3414	23	171	175	72	3855
1071	QC	QC-5	3411	25	169	174	78	3857
1144	QC	QC-7	3417	24	172	171	68	3852
1157	QC	QC-8	3403	26	170	174	75	3848
1240	ON	ON-3	3409	29	173	173	61	3845
1402	ON	ON-3	3411	24	168	172	68	3843
1409	ON	ON-5	3425	23	166	167	69	3850
1419	ON	ON-8	3421	28	171	171	64	3855
1427	QC	QC-9	3413	24	169	167	64	3837
1438	QC	QC-8	3409	23	163	173	72	3840
1452	ON	ON-2	3333	25	165	167	66	3756
1495	ON	ON-6	3420	31	170	171	62	3854
1512	ON	ON-6	3417	27	170	175	64	3853
1517	QC	QC-11	3416	24	170	176	71	3857
1578	QC	QC-12	3423	29	173	171	64	3860
K-10	N/A	NCBI	3370	0	217	177	91	3855
MAPK_JB15/16	N/A	NCBI	3379	0	185	185	76	3825
NL 89C	NL	NCBI	3327	0	186	173	68	3754
NL 93B	NL	NCBI	3327	0	183	170	66	3746
NL 95A	NL	NCBI	3337	0	181	171	65	3754
NL 95B	NL	NCBI	3332	0	181	174	66	3753
NL 95E	NL	NCBI	3309	0	181	173	63	3726
NL 96E	NL	NCBI	3325	0	184	173	66	3748
S397	N/A	NCBI	1645	0	82	116	49	1892
Telford	N/A	NCBI	0	0	0	0	0	0

Appendix Table A7: Snippy Results of MAP strains using type III strain S397 as a reference sequence

Strain ID	Province	Herd	SNPs	MNPs	INS	DEL	COMPLEX	TOTAL
23	QC	QC-1	3396	24	193	161	76	3850
32	QC	QC-1	3391	23	188	159	72	3833
117	ON	ON-1	3381	24	188	161	70	3824
137	ON	ON-1	3370	21	188	162	75	3816
160	QC	QC-2	3378	22	182	158	74	3814
171	QC	QC-3	3387	25	191	162	72	3837
183	QC	QC-2	3385	24	188	164	71	3832
191	QC	QC-2	3385	21	183	161	74	3824
206	QC	QC-3	3388	23	195	164	77	3847
211	ON	ON-2	3386	28	192	161	72	3839
219	QC	QC-1	3395	25	191	161	78	3850
233	ON	ON-1	3381	23	186	158	74	3822
236	ON	ON-1	3380	24	179	162	73	3818
242	ON	ON-1	3380	24	190	158	72	3824
252	ON	ON-1	3380	23	184	162	74	3823
325	ON	ON-3	3364	20	188	162	73	3807
326	ON	ON-3	3390	25	188	163	74	3840
342	ON	ON-4	3400	25	188	161	75	3849
345	ON	ON-4	3383	23	189	160	71	3826
366	QC	QC-1	3393	25	187	159	78	3842
375	QC	QC-4	3395	25	190	164	69	3843
392	QC	QC-5	3384	21	192	166	77	3840
400	QC	QC-4	3395	24	192	161	71	3843
418	QC	QC-3	3387	26	187	164	73	3837
441	QC	QC-6	3398	27	193	162	70	3850

464	ON	ON-5	3393	22	192	161	76	3844
500	QC	QC-7	3391	25	194	164	74	3848
505	QC	QC-7	3391	23	187	164	73	3838
508	QC	QC-8	3368	14	185	161	81	3809
515	QC	QC-9	3388	21	193	163	76	3841
516	QC	QC-9	3379	22	190	160	78	3829
531	QC	QC-10	3394	23	181	159	75	3832
552	ON	ON-2	3388	26	191	160	72	3837
560	ON	ON-2	3396	27	190	160	70	3843
586	ON	ON-6	3382	24	185	158	73	3822
594	ON	ON-6	3384	24	188	156	70	3822
631	ON	ON-7	3383	25	190	162	73	3833
638	ON	ON-7	3375	22	183	159	74	3813
648	ON	ON-7	3390	25	185	163	79	3842
863	ON	ON-4	3389	24	186	158	74	3831
869	QC	QC-4	3391	22	193	157	76	3839
878	QC	QC-4	3388	22	191	163	72	3836
885	QC	QC-5	3389	20	186	163	85	3843
897	ON	ON-5	3400	23	190	159	85	3857
910	ON	ON-8	3389	24	192	159	71	3835
918	ON	ON-8	3390	23	196	162	75	3846
928	ON	ON-8	3339	21	186	159	73	3778
938	QC	QC-10	3390	26	188	164	68	3836
948	QC	QC-10	3390	27	184	159	69	3829
1007	ON	ON-6	3338	21	185	156	76	3776
1023	ON	ON-6	3385	24	194	162	68	3833
1032	QC	QC-11	3387	25	189	166	73	3840

1044	QC	QC-11	3382	18	198	166	84	3848
1071	QC	QC-5	3387	23	191	164	81	3846
1144	QC	QC-7	3386	21	185	160	75	3827
1157	QC	QC-8	3375	24	194	164	80	3837
1240	ON	ON-3	3382	25	192	163	74	3836
1402	ON	ON-3	3380	22	189	163	68	3822
1409	ON	ON-5	3395	20	192	158	77	3842
1419	ON	ON-8	3392	23	198	159	72	3844
1427	QC	QC-9	3381	21	191	159	72	3824
1438	QC	QC-8	3383	20	191	164	82	3840
1452	ON	ON-2	3297	23	183	160	72	3735
1495	ON	ON-6	3389	28	188	160	70	3835
1512	ON	ON-6	3385	25	192	163	71	3836
1517	QC	QC-11	3389	21	193	167	80	3850
1578	QC	QC-12	3397	27	190	158	74	3846
K-10	N/A	NCBI	3349	0	241	170	96	3856
MAPK_JB15/16	N/A	NCBI	3354	0	216	180	86	3836
NL 89C	NL	NCBI	3313	0	194	167	80	3754
NL 93B	NL	NCBI	3318	0	191	163	79	3751
NL 95A	NL	NCBI	3332	0	192	164	79	3767
NL 95B	NL	NCBI	3323	0	189	167	81	3760
NL 95E	NL	NCBI	3297	0	191	169	80	3737
NL 96E	NL	NCBI	3317	0	193	166	81	3757
S397	N/A	NCBI	0	0	0	0	0	0
Telford	N/A	NCBI	1645	0	113	90	50	1898

Appendix Table A8: Snippy Results of MAP strains using type B strain MAPK_JB15/16 as a reference sequence

Strain ID	Province	Herd	SNPs	MNPs	INS	DEL	COMPLEX	TOTAL
23	QC	QC-1	645	1	51	42	9	748
32	QC	QC-1	645	2	52	42	8	749
117	ON	ON-1	633	2	50	42	7	734
137	ON	ON-1	630	1	52	43	8	734
160	QC	QC-2	640	1	53	44	10	748
171	QC	QC-3	638	2	50	45	8	743
183	QC	QC-2	640	3	54	44	6	747
191	QC	QC-2	639	3	55	45	6	748
206	QC	QC-3	646	0	52	43	10	751
211	ON	ON-2	634	4	56	42	6	742
219	QC	QC-1	647	2	51	42	8	750
233	ON	ON-1	630	3	51	44	7	735
236	ON	ON-1	629	1	53	45	10	738
242	ON	ON-1	628	3	53	42	6	732
252	ON	ON-1	629	3	50	45	8	735
325	ON	ON-3	636	1	50	43	9	739
326	ON	ON-3	643	3	51	45	8	750
342	ON	ON-4	649	4	56	43	6	758
345	ON	ON-4	639	3	48	43	8	741
366	QC	QC-1	642	3	52	40	7	744
375	QC	QC-4	641	1	52	44	7	745
392	QC	QC-5	636	2	52	48	13	751
400	QC	QC-4	641	2	55	43	6	747
418	QC	QC-3	638	3	55	47	7	750
441	QC	QC-6	640	3	55	44	6	748

464	ON	ON-5	649	1	53	43	9	755
500	QC	QC-7	647	1	55	45	9	757
505	QC	QC-7	647	2	54	46	8	757
508	QC	QC-8	634	2	53	46	12	747
515	QC	QC-9	642	2	55	45	8	752
516	QC	QC-9	639	1	56	46	9	751
531	QC	QC-10	644	3	51	46	6	750
552	ON	ON-2	631	3	54	43	7	738
560	ON	ON-2	643	3	53	42	6	747
586	ON	ON-6	641	2	53	41	7	744
594	ON	ON-6	642	1	52	44	9	748
631	ON	ON-7	642	2	51	44	10	749
638	ON	ON-7	635	4	50	43	8	740
648	ON	ON-7	639	2	53	44	10	748
863	ON	ON-4	645	1	52	43	8	749
869	QC	QC-4	642	1	53	43	7	746
878	QC	QC-4	642	1	54	44	7	748
885	QC	QC-5	641	2	50	48	12	753
897	ON	ON-5	653	1	57	43	9	763
910	ON	ON-8	649	1	53	43	9	755
918	ON	ON-8	649	2	52	43	8	754
928	ON	ON-8	590	0	48	41	9	688
938	QC	QC-10	642	2	54	44	8	750
948	QC	QC-10	641	1	51	44	8	745
1007	ON	ON-6	595	2	49	40	7	693
1023	ON	ON-6	645	2	53	43	7	750
1032	QC	QC-11	639	2	52	47	12	752

1044	QC	QC-11	644	1	53	49	13	760
1071	QC	QC-5	639	3	51	48	11	752
1144	QC	QC-7	642	3	52	47	7	751
1157	QC	QC-8	631	2	52	47	13	745
1240	ON	ON-3	643	2	50	44	10	749
1402	ON	ON-3	641	2	51	45	9	748
1409	ON	ON-5	649	3	56	43	8	759
1419	ON	ON-8	651	2	54	42	8	757
1427	QC	QC-9	639	1	50	44	9	743
1438	QC	QC-8	635	2	52	45	12	746
1452	ON	ON-2	564	3	44	41	7	659
1495	ON	ON-6	644	1	55	42	9	751
1512	ON	ON-6	646	2	51	46	8	753
1517	QC	QC-11	639	1	53	47	13	753
1578	QC	QC-12	646	1	53	43	9	752
K-10	N/A	NCBI	644	0	95	49	13	801
MAPK_JB15/16	N/A	NCBI	0	0	0	0	0	0
NL 89C	NL	NCBI	632	0	55	46	8	741
NL 93B	NL	NCBI	633	0	56	46	6	741
NL 95A	NL	NCBI	638	0	56	46	6	746
NL 95B	NL	NCBI	634	0	55	47	6	742
NL 95E	NL	NCBI	615	0	56	47	7	725
NL 96E	NL	NCBI	631	0	55	47	6	739
S397	N/A	NCBI	3355	0	173	224	85	3837
Telford	N/A	NCBI	3388	0	174	199	75	3836

Appendix Table A9: Unique variants called by Snippy using type II strain K-10 as a reference sequence

Isolate ID	Province	Herd	Unique SNPs	Unique INS	Unique DEL	Unique MNP	Unique Complex Variation
23	QC	QC-1	3	0	1	0	0
32	QC	QC-1	3	0	0	0	0
117	ON	ON-1	2	0	0	0	0
137	ON	ON-1	2	0	0	0	0
160	QC	QC-2	3	0	0	0	1
171	QC	QC-3	1	0	0	0	0
183	QC	QC-2	3	0	0	0	0
191	QC	QC-2	3	0	0	0	0
206	QC	QC-3	6	0	0	0	0
211	ON	ON-2	4	0	0	0	0
219	QC	QC-1	3	0	0	0	0
233	ON	ON-1	8	2	0	0	0
236	ON	ON-1	6	0	1	0	0
242	ON	ON-1	2	0	0	0	0
252	ON	ON-1	4	0	0	0	0
325	ON	ON-3	14	0	1	0	0
326	ON	ON-3	1	1	0	0	0
342	ON	ON-4	2	0	0	1	0
345	ON	ON-4	20	0	1	0	1
366	QC	QC-1	0	0	0	0	0
375	QC	QC-4	1	0	0	0	0
392	QC	QC-5	0	0	0	0	0
400	QC	QC-4	1	1	0	0	0
418	QC	QC-3	1	1	0	0	0
441	QC	QC-6	13	3	0	0	1

464	ON	ON-5	1	0	2	0	0
500	QC	QC-7	1	0	0	0	0
505	QC	QC-7	3	0	0	0	0
508	QC	QC-8	6	0	0	0	0
515	QC	QC-9	3	0	0	0	0
516	QC	QC-9	0	0	1	0	0
531	QC	QC-10	5	0	0	0	0
552	ON	ON-2	2	0	0	0	0
560	ON	ON-2	17	0	0	0	2
586	ON	ON-6	10	1	1	0	1
594	ON	ON-6	0	1	0	0	0
631	ON	ON-7	6	0	0	1	1
638	ON	ON-7	0	0	0	0	0
648	ON	ON-7	0	0	0	0	0
863	ON	ON-4	0	0	0	0	0
869	QC	QC-4	2	0	0	0	0
878	QC	QC-4	2	0	1	0	0
885	QC	QC-5	4	1	0	0	0
897	ON	ON-5	4	1	0	0	0
910	ON	ON-8	1	0	0	0	0
918	ON	ON-8	0	0	0	0	0
928	ON	ON-8	0	0	0	0	0
938	QC	QC-10	3	0	0	1	0
948	QC	QC-10	3	0	0	0	0
1007	ON	ON-6	2	0	0	0	0
1023	ON	ON-6	2	0	0	0	0
1032	QC	QC-11	2	0	0	0	0

1044	QC	QC-11	6	0	1	0	0
1071	QC	QC-5	0	0	0	0	0
1144	QC	QC-7	0	0	0	0	0
1157	QC	QC-8	1	0	1	0	0
1240	ON	ON-3	3	0	0	0	0
1402	ON	ON-3	0	0	0	0	0
1409	ON	ON-5	0	0	0	0	0
1419	ON	ON-8	3	0	0	0	0
1427	QC	QC-9	3	0	0	0	0
1438	QC	QC-8	1	0	0	0	0
1452	ON	ON-2	31	2	3	0	0
1495	ON	ON-6	4	0	0	0	0
1512	ON	ON-6	4	0	0	0	0
1517	QC	QC-11	1	0	0	0	0
1578	QC	QC-12	18	1	0	0	2

Appendix Table A10: Assembly statistics for MAP isolates (n=139) from Quebec and Ontario Herds

Strain ID	Animal ID	Herd ID	Genome Size (bp)	GC Content (%)	Average Coverage (X)	# Contigs	Average Contig Size	N50
#159	A19	QC-2	4780984	69.33	13	80	33325.8	118496
#160	A19	QC-2	4781128	69.33	12	83	33321.4	118496
#161	A19	QC-2	4780053	69.33	12.5	85	33309.5	118164
#162	A19	QC-2	4781121	69.33	12	76	36068.2	128709
#163	A19	QC-2	4780917	69.33	14	76	33565.8	122818
#164	A19	QC-2	4781648	69.33	14	94	31773.1	106432
#165	A19	QC-2	4781476	69.33	16	87	30776.1	106432
#166	A19	QC-2	4780467	69.33	18	83	27468.7	118436
#167	A19	QC-2	4781432	69.33	13	75	34774	122818
#168	A19	QC-2	4781342	69.33	14	73	35030	131207
#169	A20	QC-3	4780607	69.33	14	77	31592.9	122816
#170	A20	QC-3	4780884	69.33	12	76	36613	122818
#171	A20	QC-3	4781233	69.33	16	73	30997.6	131207
#172	A20	QC-3	4781204	69.33	17	73	33809.1	128709
#173	A20	QC-3	4781259	69.33	17	73	32020.7	131207
#174	A20	QC-3	4781758	69.33	15	77	32227	118496
#175	A20	QC-3	4780836	69.33	16	76	30599.8	124823
#176	A20	QC-3	4781907	69.33	18.5	78	30031.9	122816
#177	A20	QC-3	4781301	69.33	12	77	36893	118496
#179	A21	QC-2	4781745	69.33	16	81	31798	118164
#180	A21	QC-2	4781402	69.33	14	74	35546.6	131207
#181	A21	QC-2	4781851	69.33	15.5	84	35266.3	118120
#182	A21	QC-2	4780491	69.33	12	78	35790.2	122818
#183	A21	QC-2	4780754	69.33	13	78	34271.9	128709
#184	A21	QC-2	4782404	69.33	15	79	37763.8	118496

#185	A21	QC-2	4781501	69.33	16	73	34783.5	122818
#186	A21	QC-2	4781081	69.33	15	74	32885.6	128709
#187	A21	QC-2	4781464	69.33	16	77	31801.6	118496
#188	A21	QC-2	4781466	69.33	19	83	27165.4	118164
#189	A22	QC-2	4781233	69.33	16	84	33783.8	118164
#190	A22	QC-2	4781273	69.33	13	87	35254.8	118043
#191	A22	QC-2	4781072	69.33	16	85	31784.2	118164
#192	A22	QC-2	4781509	69.33	10	87	39590.5	112918
#193	A22	QC-2	4781719	69.33	14	83	33557.9	118059
#194	A22	QC-2	4781229	69.33	15	81	32652.7	118496
#195	A22	QC-2	4782076	69.33	11	73	38990.6	131207
#196	A22	QC-2	4780838	69.33	14	82	35786	118164
#197	A22	QC-2	4780947	69.33	14	82	34023	118436
#198	A22	QC-2	4781369	69.33	17	84	29292.2	118057
#199	A23	QC-3	4780905	69.33	18.5	84	30973.1	118149
#200	A23	QC-3	4781711	69.33	13	86	35785.3	118072
#201	A23	QC-3	4781774	69.33	15	86	33320.3	118137
#202	A23	QC-3	4780787	69.33	12	80	36882.7	118436
#203	A23	QC-3	4781872	69.33	16	74	32022.8	128709
#204	A23	QC-3	4780660	69.33	13	75	34276.6	131207
#205	A23	QC-3	4780922	69.33	14	73	35543.9	131207
#206	A23	QC-3	4781381	69.33	16.5	77	32442.6	122818
#207	A23	QC-3	4780812	69.33	16	73	32018.9	131207
#208	A23	QC-3	4780995	69.33	14.5	75	35026.5	131207
#212	A25	QC-1	4780772	69.33	14	83	33784.3	118495
#213	A25	QC-1	4780619	69.33	12	84	37155.5	118164
#23	A25	QC-1	4781165	69.33	18	79	32434.6	118495
#24	A25	QC-1	4781113	69.33	17	89	30966.6	106432

#25	A25	QC-1	4780757	69.33	14.5	80	32217.3	118164
#26	A25	QC-1	4779226	69.33	12	83	34745	118164
#27	A25	QC-1	4780190	69.33	13	84	32861.3	118495
#28	A25	QC-1	4781527	69.33	14	88	34256.9	114137
#29	A25	QC-1	4780219	69.33	12	77	37170.5	118495
#30	A25	QC-1	4780360	69.33	11	76	39608.1	124653
#214	A26	QC-1	4780462	69.33	12	81	36598.8	118129
#215	A26	QC-1	4780251	69.33	11	84	37731.6	125592
#216	A26	QC-1	4780417	69.33	12	85	33547.3	118720
#32	A26	QC-1	4779851	69.33	13	81	34752.6	118495
#33	A26	QC-1	4780983	69.33	15	81	32215.5	118495
#34	A26	QC-1	4780748	69.33	19.5	88	28257.6	102083
#35	A26	QC-1	4779720	69.33	12	83	33540.6	118164
#36	A26	QC-1	4780604	69.33	14	78	33099.3	122816
#37	A26	QC-1	4780494	69.33	14	79	34761.4	118495
#38	A26	QC-1	4781354	69.33	14	74	35807.2	131207
#113	A32	ON-1	4781895	69.33	13	77	34526.2	128709
#114	A32	ON-1	4782036	69.33	12	81	35796.2	118163
#115	A32	ON-1	4781342	69.33	15	79	30401.9	122816
#116	A32	ON-1	4780709	69.33	14.5	80	30980.4	118498
#117	A32	ON-1	4780632	69.33	16	83	29830.5	128709
#118	A32	ON-1	4779602	69.33	14	81	32205.9	118498
#119	A32	ON-1	4780836	69.33	16.5	86	29284.9	106432
#120	A32	ON-1	4780613	69.33	12	91	34001.1	102076
#121	A32	ON-1	4781430	69.33	18	82	27787.6	118163
#122	A32	ON-1	4779567	69.33	14	86	32850.8	106432
#126	A34	ON-1	4779715	69.33	14.5	89	29272.1	113042
#127	A34	ON-1	4780569	69.33	11	86	34254.8	118164

#128	A34	ON-1	4780440	69.33	11	86	33309.5	118055
#129	A34	ON-1	4780008	69.33	14.5	90	29271.4	106432
#130	A34	ON-1	4780968	69.33	15	88	33309.3	113042
#131	A34	ON-1	4780491	69.33	13.5	82	36320.5	118495
#132	A34	ON-1	4780404	69.33	15	87	28423.9	118164
#133	A34	ON-1	4781072	69.33	11.5	89	34216.3	106432
#236	A34	ON-1	4780673	69.33	12	89	30377.5	106432
#237	A34	ON-1	4780812	69.33	15	81	32002.4	118495
#136	A36	ON-1	4780657	69.33	14.5	77	31592.6	125592
#247	A36	ON-1	4780585	69.33	11	77	37459.8	125592
#248	A36	ON-1	4779666	69.33	14.5	86	31362.4	113042
#249	A36	ON-1	4780957	69.33	16	78	30793.3	118495
#250	A36	ON-1	4780983	69.33	12	83	34021.8	118164
#251	A36	ON-1	4781085	69.33	10.5	82	37740.4	118163
#252	A36	ON-1	4780653	69.33	13	80	36331.1	118164
#253	A36	ON-1	4780066	69.33	12	79	35262.8	118498
#254	A36	ON-1	4780612	69.33	10	80	37167.4	125592
#255	A36	ON-1	4780353	69.33	10	79	38967.2	118495
#137	A37	ON-1	4779052	69.33	14	78	33780.2	118498
#256	A37	ON-1	4778417	69.33	13	82	33075.9	118163
#258	A37	ON-1	4778665	69.33	12	77	35516.2	122818
#259	A37	ON-1	4778378	69.33	12	84	33998	118163
#260	A37	ON-1	4780723	69.33	15	82	34020.2	118498
#261	A37	ON-1	4781526	69.33	14	82	31998.6	122816
#262	A37	ON-1	4780254	69.33	13	83	36046	118498
#263	A37	ON-1	4778854	69.33	13	84	32191.8	106432
#264	A37	ON-1	4779724	69.33	12	89	35240.5	101893
#265	A37	ON-1	4778750	69.33	13	82	33532.6	118163

#217	AM1	QC-1	4780603	69.33	14	76	32013.1	122816
#218	AM1	QC-1	4779441	69.33	11.5	84	34495.4	106432
#219	AM1	QC-1	4781551	69.33	14.5	76	36069.7	131207
#220	AM1	QC-1	4781103	69.33	14	74	37469.4	124649
#221	AM1	QC-1	4780343	69.33	15	83	31576.9	118046
#222	AM1	QC-1	4781285	69.33	14	88	32419.9	99341
#223	AM1	QC-1	4781657	69.33	14.5	77	34771.5	128709
#224	AM1	QC-1	4781293	69.33	16	76	35025.3	122816
#225	AM1	QC-1	4780208	69.33	15	81	31177.4	118435
#226	AM1	QC-1	4781279	69.33	14	79	32878.6	122816
#123	AM2	ON-1	4780641	69.33	14	95	31559.6	99322
#124	AM2	ON-1	4780602	69.33	13	94	32627.3	102071
#125	AM2	ON-1	4780107	69.33	15	86	30191.7	113042
#228	AM2	ON-1	4780310	69.33	15	82	30201.1	118164
#229	AM2	ON-1	4780355	69.33	16	79	28774.9	118164
#230	AM2	ON-1	4780646	69.33	16	94	30562.7	113042
#231	AM2	ON-1	4781165	69.33	18	79	32434.6	118495
#232	AM2	ON-1	4780376	69.33	14.5	92	30189.3	99337
#233	AM2	ON-1	4780255	69.33	13	86	34009.7	118071
#234	AM2	ON-1	4780307	69.33	15.5	87	30965.5	118164
#134	AM3	ON-1	4780112	69.33	13	77	35527.4	122816
#135	AM3	ON-1	4781400	69.33	14	80	35796.7	118495
#238	AM3	ON-1	4780676	69.33	12	83	31785.1	118163
#239	AM3	ON-1	4780114	69.33	14	82	34500.5	118495
#240	AM3	ON-1	4779719	69.33	13	77	37161.4	131207
#241	AM3	ON-1	4780663	69.33	13	77	35272.1	131207
#242	AM3	ON-1	4780988	69.33	16	74	27639.3	131207
#243	AM3	ON-1	4781130	69.33	13	78	33564.1	125592

#244	AM3	ON-1	4780531	69.33	13.5	79	34511.6	118498
#245	AM3	ON-1	4780153	69.33	14	76	34516.2	131207
	Average		4780751.22	69.33	14.03	81.22	33438.86	119229.19

Appendix Table A11: CheckM Results to Determine Assembly Completeness and Contamination in MAP isolates (n=139) from Quebec and Ontario Herds

Isolate Number	Farm ID	Animal Number	Marker Lineage	# Genomes	# Markers	# Marker Sets	0	1	2	3	4	5+	Completeness (%)	Contamination (%)
#159	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.42
#160	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.42
#161	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4
#162	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.49
#163	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#164	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#165	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.57
#166	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.39
#167	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#168	QC-2	A19	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.64
#169	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#170	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#171	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#172	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.88
#173	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.78
#174	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4
#175	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#176	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#177	QC-3	A20	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.42
#179	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33

#180	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.61
#181	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.51
#182	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#183	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.67
#184	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.54
#185	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	684	6	0	0	0	99.28	1.25
#186	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.67
#187	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.71
#188	QC-2	A21	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.49
#189	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.42
#190	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.43
#191	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#192	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#193	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4
#194	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	0.82
#195	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.36
#196	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#197	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	2	688	3	0	0	0	99.61	0.7
#198	QC-2	A22	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.51
#199	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.67
#200	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.51
#201	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5
#202	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5

#203	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	684	6	0	0	0	99.28	1.17
#204	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.46
#205	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#206	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#207	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	4	0	0	0	99.28	0.47
#208	QC-3	A23	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5
#212	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	693	3	0	0	0	99.28	0.46
#213	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.3
#23	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4
#24	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.47
#25	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.38
#26	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#27	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#28	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.5
#29	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.78
#30	QC-1	A25	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.66
#214	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#215	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#216	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.44
#32	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.39
#33	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.86
#34	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#35	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.72

#36	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#37	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5
#38	QC-1	A26	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#113	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.5
#114	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	2	690	1	0	0	0	99.61	0.33
#115	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.89
#116	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.58
#117	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	1.11
#118	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.73
#119	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.53
#120	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#121	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.67
#122	ON-1	A32	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#126	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.89
#127	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#128	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.39
#129	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	0.69
#130	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#131	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#132	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.78
#133	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#236	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	1.28
#237	ON-1	A34	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.75

#136	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.36
#247	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#248	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.58
#249	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.57
#250	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.5
#251	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5
#252	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#253	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#254	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#255	ON-1	A36	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#137	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.36
#256	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.67
#258	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.47
#259	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#260	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#261	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4
#262	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4
#263	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#264	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.42
#265	ON-1	A37	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.61
#217	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#218	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#219	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.88

#220	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#221	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	1
#222	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.44
#223	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#224	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.53
#225	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#226	QC-1	AM1	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.42
#123	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	688	3	0	0	0	99.28	0.67
#124	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.47
#125	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	688	3	0	0	0	99.28	0.67
#228	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	688	3	0	0	0	99.28	0.5
#229	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	689	3	0	0	0	99.28	0.33
#230	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	689	3	0	0	0	99.28	0.33
#231	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	688	3	0	0	0	99.28	0.4
#232	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	686	3	0	0	0	99.28	0.48
#233	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	689	3	0	0	0	99.28	0.33
#234	ON-1	AM2	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.67
#134	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.58
#135	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33
#238	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	686	4	0	0	0	99.28	0.55
#239	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.36
#240	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.5
#241	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	689	1	0	0	0	99.28	0.33

#242	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	685	5	0	0	0	99.28	0.85
#243	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	687	3	0	0	0	99.28	0.44
#244	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.36
#245	ON-1	AM3	g__Mycobacterium (UID1816)	100	693	300	3	688	2	0	0	0	99.28	0.4

Appendix Table A12: Homologous *Mycobacterium avium* complex nucleotide sequences of ORF-1 and ORF-2 identified by BLASTn

SSR Examined	*Mycobacterium* species	Strain Name	Sequence ID (NCBI)	Nucleotide Sequence
ORF-1	*Mycobacterium avium* subsp. *avium*	104	CP000479.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAAGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACGATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCAGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACGCTGTATCAGCTGCCCAACGACTATCAGCTGCCC GCCGGCCTGGCGGGGGTGCAGGCGGCGCAGACGATCGGCG CCGCCGCGACCGCGCAGGCTCAGCTCACCGGGGTGTTCGG CGGCTGA
ORF-1	*Mycobacterium avium* subsp. *avium*	DSM 44156	CP046507.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAGGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACAATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACCCTCTATCAGCTGCCCAACGACTATCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGGCCATTGCCG GCGCCGCGACCGCGCAGGCGCAGCTCACCGGGGTGTTCGG CGGCTGA

ORF-1	*Mycobacterium avium* subsp. *avium*	RCAD0278	CP016396.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAGGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACAATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACCCTCTATCAGCTGCCCAACGACTATCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGGCCATTGCCG GCGCCGCGACCGCGCAGGCGCAGCTCACCGGGGTGTTCGG CGGCTGA
ORF-1	*Mycobacterium avium* subsp. *avium*	FDAARGOS_1609	CP089223.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAGGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACAATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACCCTCTATCAGCTGCCCAACGACTATCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGGCCATTGCCG GCGCCGCGACCGCGCAGGCGCAGCTCACCGGGGTGTTCGG CGGCTGA

ORF-1	*Mycobacterium avium* subsp. *avium*	FDAARGOS_1608	CP089230.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAGGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACAATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACCCTCTATCAGCTGCCCAACGACTATCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGGCCATTGCCG GCGCCGCGACCGCGCAGGCGCAGCTCACCGGGGTGTTCGG CGGCTGA
ORF-1	*Mycobacterium avium* subsp. *hominissuis*	101115	CP040255.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAAGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACGATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCAGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACGCTGTATCAGCTGCCCAACGACTATCAGCTGCCC GCCGGCCTGGCGGGGGTGCAGGCGGCGCAGACGATCGGCG CCGCCGCGACCGCGCAGGCTCAGCTCACCGGGGTGTTCGG CGGCTGA

ORF-1	*Mycobacterium avium* subsp. *hominissuis*	101034	CP040247.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAAGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACGATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCAGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACGCTGTATCAGCTGCCCAACGACTATCAGCTGCCC GCCGGCCTGGCGGGGGTGCAGGCGGCGCAGACGATCGGCG CCGCCGCGACCGCGCAGGCTCAGCTCACCGGGGTGTTCGG CGGCTGA
ORF-1	*Mycobacterium avium* subsp. *hominissuis*	101174	CP040250.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAAGGTGACCACCCTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACGATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGCCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCAGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACGCTGTATCAGCTGCCCAACGACTATCAGCTGCCC GCCGGCCTGGCGGGGGTGCAGGCGGCGCAGACGATCGGCG CCGCCGCGACCGCGCAGGCTCAGCTCACCGGGGTGTTCGG CGGCTGA

ORF-1	*Mycobacterium avium* subsp. *hominissuis*	JP-H-1	AP020326.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAAGGTGACCACCCTGGGCCGCACGGCGCAGAACGC CCCGGCCGGCGGACCGTTGACGATCACCAGCCCGGCGTTC ACCGACGGCGCCCCGATCCCGGCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACCCTCTATCAGCTGCCCAACGACTACCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGACCATTGCCG GCGCCGCCACCGCGCAGGCGCAGCTCACCGGGACGTTCGG CGGCTGA
ORF-1	*Mycobacterium avium* subsp. *hominissuis*	MAH11	CP035744.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCA CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCGGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAAGGTGACCACCCTGGGCCGCACGGCGCAGAACGC CCCGGCCGGCGGACCGTTGACGATCACCAGCCCGGCGTTC ACCGACGGCGCCCCGATCCCGGCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCGCTGGCGTGGTCGGCGCCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCCGGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACGCCGCCCGGCAC AACGACTTTGCCGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACCCTCTATCAGCTGCCCAACGACTACCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGACCATTGCCG GCGCCGCCACCGCGCAGGCGCAGCTCACCGGGACGTTCGG CGGCTGA

ORF-1	*Mycobacterium avium subsp. lepraemurium*	Hawaii	CP021238.1	ATGGAAGCGACGCTGGTGGACACCTCTCGCCGGATCGGCG CGATCATCTGCGCGCTGGCGCTGCCGGCGGGCGCAGTCGGT TGCGGCGGCCACGGTCATGGCCCGACGACGACTCCGTCGA CACCCAGGGTGACCACCTTGGGCCGCACGGCGCCGAACGC CCCGGCCGGCGGGCCGTTGACGATCACCAGCCCGGCATTC ACCGACGGCGCCCCGATCCCGGCGCGGTACACCTGCAAGG GCGAGGGCATCGCACCGCCACTGGCGTGGTCCGCACCGAC GGGTGCGGCCCTGGTCGTCGACGACCCCGACGCCCTAGCG GGGCCGTACGTGCACTGGGTGGTGACCGGCATCGCCCCGG GCTCCGGCAGCACGTCCGCGGGCCAAACACCGCCCGGCAC AACGACTTTGCAGAACACCGCCGGACAGGCCGGCTACCAG GGGCCGTGTCCGCCCGCCGGAACGGGCACCCACCACTACC GGTTCACGCTGTATCGGCTGCCCAACGACTACCAGCTGCCC GGTGGGCTGGCCGGTGCGCAGGCCGCGCAGACCATTGCCG GCGCCGCCACCGCGCAGGCGCAGCTCACCGGGACGTTCGG CGGCTGA
ORF-2	*Mycobacterium avium* subsp. *avium*	FDAARGOS_1609	CP089223.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGGGCTTCGGTGACGTCGAAGCTCACCAC AGGTCTGCCGAGCGGCATCGAGCGGCACGGGATGCGGACT ACCTGGCTGCTGAGCAGGATCAATAAAGCCTGA
ORF-2	*Mycobacterium avium* subsp. *avium*	104	CP000479.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGGGCTTCGGTGA

ORF-2	*Mycobacterium avium* subsp. *avium*	FDAARGOS_1608	CP089230.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTTGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGGGCTTCGGTGACGTCGAAGCTCACCAC AGGTCTGCCGAGCGGCATCGAGCGGCACGGGATGCGGACT ACCTGGCTGCTGAGCAGGATCAATAAAGCCTGA
ORF-2	*Mycobacterium avium* subsp. *avium*	DSM 44156	CP046507.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGCTTCGGTGACGTCGAAGCTCACCACAG GTCTGCCGAGCGGCATCGAGCGGCACGGGATGCGGACTAC CTGGCTGCTGAGCAGGATCAATAA
ORF-2	*Mycobacterium avium* subsp. *avium*	HJW	CP028731.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGCTTCGGTGACGTCGAAGCTCACCACAG GTCTGCCGAGCGGCATCGAGCGGCACGGGATGCGGACTAC CTGGCTGCTGAGCAGGATCAATAA
ORF-2	*Mycobacterium avium* subsp. *hominissuis*	MAC109	CP029332.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGTGAACGTCTGCATGCGGCCCGCCTGTCCGCCG GCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAGC TGCTGGGGGGGGGGGCTTCGGTGACGTCGAAGCCCACCAC AGGTCTGCCGAGCGGCATCGAGCGGCACGGGATGCGGACT ACCTGGCTGCTGAGCAGGATCAGTAA
ORF-2	*Mycobacterium avium* subsp. *hominissuis*	W9	CP060405.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGTGAACGTCTGCATGCGGCCCGCCTGTCCGCCG GCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAGC TGCTGGGGGGGGGCTTCGGTGA

ORF-2	*Mycobacterium avium* subsp. *hominissuis*	H87	CP018363.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGGGCTTCGGTGACGTCGAAGCTCACCAC AGGTCTGCCGAGCGGCATCGAGCGGCACGGGATGCGGACT ACCTGGCTGCTGAGCAGGATCAGTAAAGCCTGA
ORF-2	*Mycobacterium avium* subsp. *hominissuis*	CAM177	CP076851.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGTGAACGTCTGCATGCGGCCCGCCTGTCCGCCG GCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAGC TGCTGGGGGGGGGCTTCGGTGA
ORF-2	*Mycobacterium avium* subsp. *hominissuis*	JP-H-1	AP020326.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTGTCGC GGCGGCACGCGAACGTCTGCATGCGGCCCGCCTGTCCGCC GGCGATGCCCATGAGCGTGCCGCGCAGTTGCACGACAGAG CTGCTGGGGGGGGGCTTCGGTGA
ORF-2	*Mycobacterium avium subsp. lepraemurium*	Hawaii	CP021238.1	ATGAGCGACGACGCCGAACGGCGGTCCAGAGTTTCGGGAG CCGCACGTAGGTCCGCTGATGAGGCGGCCGAGCGTATCGC GGCGGCACGTGAACGTCTGCATGCGGCCCGCCTGTCCACCG GCGATGCCCCTGAGCGTGCCGCGGAGTTGCACGACAGAGC TGCTGGGGGGGGGCTTCGGTGA

Appendix III: Fragment Analysis and Multiplex Assay for SSR Identification, an Unpublished Chapter

Appendix 3.1: Introduction

A previous study examined ML-SSR typing of MAP on 85 isolates grown from 18 animals from Newfoundland farms [1]. At the time, currently available techniques had difficulty in analyzing mononucleotide repeats, such as those found in loci SSR1 and SSR2; only able to accurately type repeats up to 15 nucleotides long through a mass spectrometry-based approach [2]. The purpose of this study was to develop a method that allowed for cheap, reliable, and reproducible sequencing for SSR repeats beyond the 15-nucleotide limit. Work by Podder et al. [1], used a fragment analysis-based approach to analyze the repeat sizes at ML-SSR loci SSR1, SSR2, SSR8 and SSR9, with SSR1 and SSR2 being mononucleotide repeats and SSR8 and SSR9 being trinucleotide repeats [3].

The repeat lengths of the SSR loci in MAP strain K-10 have previously been described using WGS-based methods (SSR1: 19 G's, SSR2: 10 G's, SSR8: 5 repeats of "GGT", SSR9: 5 repeats of "TGC"), allowing for its use as a standard to identify the repeat sizes of other isolates [4,5]. A maximum size of 21 mononucleotide repeats was able to be detected in SSR1 using fragment analysis [1]. Analysis of 68 viable isolates identified 40 distinct SSR types, with multiple types detected in all 18 animals, with some strains detected between farms suggesting inter-herd transmission or a common source of infection [1]. At the time of the study, the confirmation of multiple MAP SSR types from the same animal seemed to be indicative of either microevolution of the unstable repeat elements found at SSR loci or the presence of true MSI events within the same animal. To test this, a follow-up study examined six MAP isolates (three from different animals

across different farms, and three derived from the same animal) using WGS technology to determine whether the presence of multiple MAP SSR types was due to microevolution or an MSI event [6]. A comparison of the number of SNPs identified within each strain showed that the number of divergent SNPs extended beyond the expected number that would be present based on the molecular clock of MAP, supporting the MSI hypothesis.

A limiting factor in the confirmation of MAP MSIs is the extensive time required to identify MAP-infected animals, decontaminate their respective fecal samples, and grow and extract the DNA from multiple colonies derived from a single animal. To address this limitation, two PCR-based assays were developed and utilized to determine if reliable identification of MAP MSIs could be obtained through direct analysis of fecal samples. Both the fragment analysis-based approach on four SSR loci as described in previous studies [1,7], alongside an amplicon sequencing-based method on SSR1 were examined.

Appendix 3.2: Materials and Methods

Appendix 3.2.1: Sample Collection

To test the feasibility of a multiplex-based approach analyzing SSR1 and SSR2, DNA was extracted at three different points within the culturing process; once from unprocessed fecal samples, once after decontamination and growth on solid media, and once after axenic cultures were grown in liquid media **(Appendix Figure A6)**. This was done with the intent that assay optimization would be first performed using axenic cultures to determine initial feasibility. If successful, the assay would then be tested on DNA extracts from a collection of multiple strains (lawn DNA) and unprocessed feces

(fecal DNA), to determine its ability to differentiate between multiple strains of MAP and in heavily contaminated environments, respectively.

The presence of MAP within each fecal sample was confirmed using the VetMAX™-Gold MAP Detection Kit (Life Technologies, Corp., Austin, TX) and all DNA extractions were performed using the ZR-96 Fecal DNA Kit (Zymo Research Corp., Irvine, CA). Once fecal samples were confirmed as being MAP positive, DNA extraction was performed using a portion of the unprocessed fecal sample. This DNA extract, referred to as fecal DNA, contains not only DNA for MAP but also the DNA of other microbes present within the fecal sample. Fecal samples were then decontaminated, and colonies of MAP were grown on solid media as described in **Figures 2.2** and **2.3**. A number of these colonies were selected for further growth within liquid media as axenic cultures. All remaining colonies on solid media were mixed, and the DNA of the collective mixture was extracted, referred to as lawn DNA. Lawn DNA extracts, due to the presence of multiple colonies of MAP, potentially contain multiple strains of MAP. A third set of DNA extractions was performed on each liquid media-grown axenic culture, with each extraction representative of a single strain of MA, referred to as axenic DNA.

Appendix 3.2.2: PCR Primer Design for Fragment Analysis

PCR primers for SSR loci 2, 8 and 9 were the same as described in Podder et al. [1]. A new primer set for SSR1 was developed using Primer-BLAST and Primer3Plus [8,9]. All PCRs were performed using Phusion® High Fidelity PCR Mastermix with GC Buffer (New England Biolabs Inc.) and 3% dimethyl sulfoxide (DMSO). Initial validation of each PCR protocol (**Appendix Table A13)** was performed with gel electrophoresis of

1.5% agarose gels. Primer statistics **(Appendix Table A14)** were calculated using OligoCalc [10].

Appendix 3.2.3: Fragment Analysis

Once primer specificity was confirmed using reference strain K-10, a primer from each pair was labelled at the 5' end with 6-carboxy fluorescein (6-FAM), and the PCR for all four primer pairs was repeated with the fluorescent primer. Validation of the fluorescent PCR was performed through gel electrophoresis of 1.5% agarose gels. Labelled PCR products were purified using the EZ-10 Spin Column PCR Products Purification Kit (Bio Basic). Due to the presence of the 6-FAM label, DNA was unable to be quantified through spectrophotometry-based methods. Instead, a portion of purified PCR products was examined using gel electrophoresis, and concentrations were diluted based on band intensity: products of high concentration resulted in the oversaturation of peaks and decreased quality of fragment analysis results. Examination of 139 axenic DNA extracts originating from 14 cattle, along with DNA from K-10 were analyzed using the primer pairs specific to SSR1 and SSR2. Fragment analysis results were compared to the expected sizes for the same strains, as determined through NovaSeq 6000 WGS results described in Chapter V of the main text. Purified PCR products were sent to The Centre for Applied Genomics (TCAG) to undergo genetic analysis using the Applied Biosystems 3730xl or 3130 capillary electrophoresis instruments. All samples were analyzed along with a GeneScan™ 500 LIZ™ dye size standard (Applied Biosystems), allowing for the detection of fragments from 35-500 nucleotides long. Fragment analysis

peaks were analyzed using the Peak Scanner tool in the Thermo Fisher Connect™ platform. The assay was repeated until consistent results were obtained for each sample.

Appendix 3.2.4: Amplicon Sequencing

To prepare samples for amplicon sequencing, two PCR reactions were performed, one to target and amplify the SSR target of interest, and the other to add the necessary indexing tags required for sample identification when sequencing **(Appendix Figure A6)**. Primers designed for target amplification consisted of a portion designed to bind and amplify their target sequence, along with a 3' tailed portion designed to bind to the indexing primers **(Appendix Table A15)**. Both indexing primers and the complementary 3' portions were provided by Dr. Brian Boyle of Laval University's Genomic Analysis Platform **(Appendix Table A16)**. All PCR reactions were performed using Q5® High-Fidelity DNA Polymerase (New England Biolabs Inc.), with Q5® GC Enhancer added to each mixture.

Amplification PCRs were performed using 10 DNA extracts from three animals (n=30) previously examined using NovaSeq 6000 whole genome sequencing technology. Validation of amplicons was performed using gel electrophoresis with 1.5% agarose gels. To pool samples from the same animal in approximately equimolar proportions, band intensities were compared using ImageJ [11], and the required amount of each sample to be pooled was calculated. Purification of both pooled PCR products and indexing PCR products was performed using the EZ-10 Spin Column PCR Products Purification Kit (Bio Basic). In addition to the axenic mixtures, lawn DNA extracts from the same animals were also processed to ensure that any repeat sizes not recorded in previous work could

still be detected. Both axenic mixtures and lawn DNA, after amplification, indexing and purification, were sent to Laval University's Genomic Analysis Platform for amplification sequencing using MiSeq technology.

Reads produced by amplification sequencing were assessed using Fastp [12] to trim adaptors from FASTQ files and remove any unpaired reads. To extract intact SSR repeats from both forward and reverse reads, the Linux "grep" command was used to extract and count the number of reads that mononucleotide repeats at intervals between 7 and 21 nucleotides. For axenic mixtures, read counts of each repeat size were compared with the previously detected sizes of repeats of each isolate used in the axenic mixture to determine the accuracy of the assay.

Appendix 3.3: Result and Discussion

Appendix 3.3.1: Initial Optimization of Fragment Analysis Using K-10

To determine the validity of the fragment analysis approach, fluorescent primers specific for SSR loci 1, 2, 8 and 9 were tested using MAP reference strain K-10 (NC_002944.2). Initial fragment analysis results showed that peaks for all four loci were slightly smaller than the expected PCR product size **(Appendix Table A17)**, an observation which was noted on each repeated fragment analysis attempt. The electrophoretic mobility of DNA fragments may be altered by many factors, including buffer type, concentration, sequence and secondary or tertiary structure formation [13]. If consistent results were obtained between electrophoresis runs, peak sizes are considered accurate despite the slight deviation observed between fragment size and expected PCR product size. While fragment analysis results for SSR2, SSR8 and SSR9 all showed

individual peaks with consistent sizes between runs, the results for SSR1 showed multiple peaks very close together **(Appendix Figure A7)**. In these cases, a single major peak was usually obvious, with smaller shoulder peaks on either side, suggesting the presence of at least three different fragment sizes, each differing by one nucleotide. These results were only observed for SSR1 and may be due to stuttering of the polymerase. Long mononucleotide repeats of G, such as those observed in SSR1 cause enzymes within PCR reactions to undergo slippage, especially for repeat sizes greater than 15 nucleotides when using some fusion enzymes [14,15]. This stuttering was observed across multiple capillary electrophoresis runs of SSR1 of K-10, with the largest peak always around 292 bp. For this work, these results were accepted, and the K-10 fragment analysis sizes were used as a benchmark to base all other results around.

Appendix 3.3.2: Fragment Analysis of Axenic DNA Extracts

As noted within prior studies [315,691], the trinucleotide repeats in SSR8 and SSR9 are not as discriminatory as the mononucleotide repeats in SSR1 or SSR2. This was especially true for the 139 axenic cultures examined in this study, as previously discussed in Chapter V and other works that used next-generation sequencing approaches, which showed significant discriminatory ability mainly within SSR1 and SSR2. As such, fragment analysis was only performed on DNA from the same 139 strains described in Chapter V using SSR1 and SSR2 primers. Fragment analysis was repeated for these strains until consistent fragment results could be observed, though in some cases and as previously observed in the K-10 SSR1 results, multiple fragments were present. By calculating the difference between the observed fragments and the expected results as

observed for MAP K-10, the expected mononucleotide repeat size for each strain could be determined, except in cases where either a definitive fragment analysis result or repeat size from WGS could not be obtained **(Appendix Table A18)**. From these results, it was noted that none of the fragment analysis results obtained for SSR1 matched those identified by whole genome sequencing. All recorded fragment analysis results for SSR1 showed higher sizes than those obtained by WGS, with many isolates showing the presence of multiple peaks in some cases. The fragment analysis and WGS results for SSR2 matched better than those of SSR1 in all cases, with results from A25, A26, A34, A37, and A19 matching perfectly for this repeat. While a portion of the isolates diverged from the expected results generated by WGS, the differences were generally closer than those observed for SSR1. While the exact reason for this is uncertain, the longer mononucleotide sequences present in SSR1 may cause slippage in the polymerase during the PCR step, as noted in the literature [689,690]. In contrast, the repeat sizes for SSR2 were generally smaller than those in SSR1, with most isolates having repeats 9-12 bp in length. These shorter repeats would limit the amount of slippage shown and would enhance the accuracy of fragment analysis in these cases. Due to the considerable amount of error shown using this method, further optimization of the assay, along with the use of lawn and fecal DNA, was stopped.

Appendix 3.3.3: Amplicon Sequencing

With the failure of fragment analysis, a different approach was taken to determine if a PCR-based method could be eventually applied to fecal DNA to identify potential MSI events. To account for the slippage previously observed during fragment analysis, an

amplicon sequencing approach was taken to target the homopolymeric repeat of SSR1. The amplicon sequencing approach developed used two PCRs, one to amplify the target using the modified primers shown in **Appendix Table A15**, and a second to tag the amplicons using the indexing primers shown in **Appendix Table A16**, allowing for the amplicon to be directly amplified and sequenced using next generation sequencing methods. Testing of the assay was performed using available axenic and lawn DNA from animals A23, A25 and A26. The amplicon sequencing assay was tested using two amplification primer sets: C (SSR1c-III-F and SSR1c-III-R) and D (SSR1d-III-F and SSR1d-III-R) **(Appendix Table A8)**.

Using pools of axenic DNA from A23, A25, and A26, the assay was able to successfully target, amplify and sequence the SSR1 repeat sequence. However, examination of the reads produced using both primer sets C and D shows that this method also lacked specificity. Using the previously established SSR repeat sizes attained by WGS as a standard, it was found that several reads produced from the results of the axenic pool analysis either contained the incorrect repeat sizes, producing reads for repeat sizes that were not recorded as being in the pooled mixture, or were missing the repeat sizes that were expected to be in the mixture **(Appendix Table A19)**. It is possible that the initial PCR also resulted in frameshift mutations which altered the number of repeats for some reads, as in fragment analysis, despite the use of a less error-prone polymerase. It is also possible that the use of the more error-prone MiSeq sequencing, instead of NovaSeq as in the WGS analysis, resulted in the sequencing errors observed [17]. While results were also obtained for the lawn DNA of these three animals **(Appendix Table**

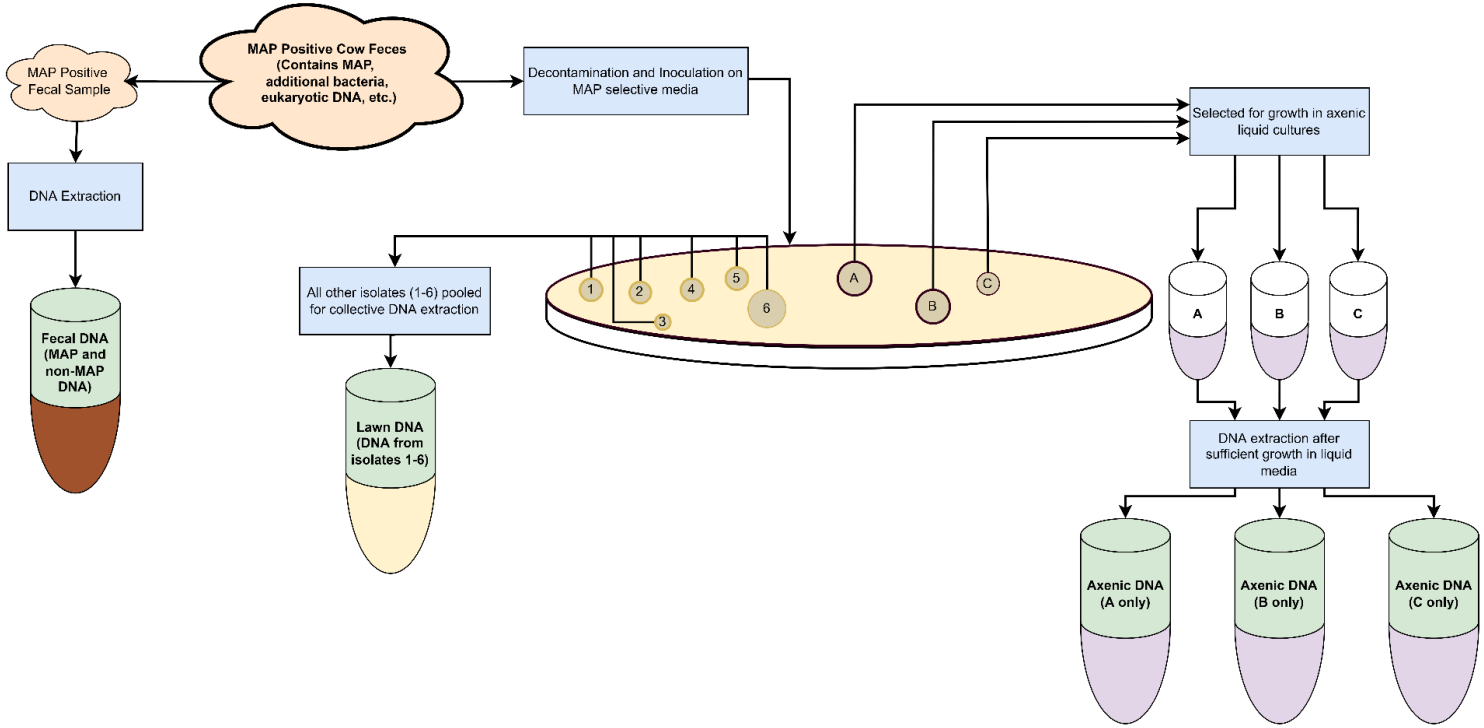
A20), due to the errors observed in the axenic mixtures, it is not certain that the variety of reads on display within the lawns is representative of any SSR1 repeat sizes that were not captured during WGS analysis.

Appendix 3.4: Conclusion

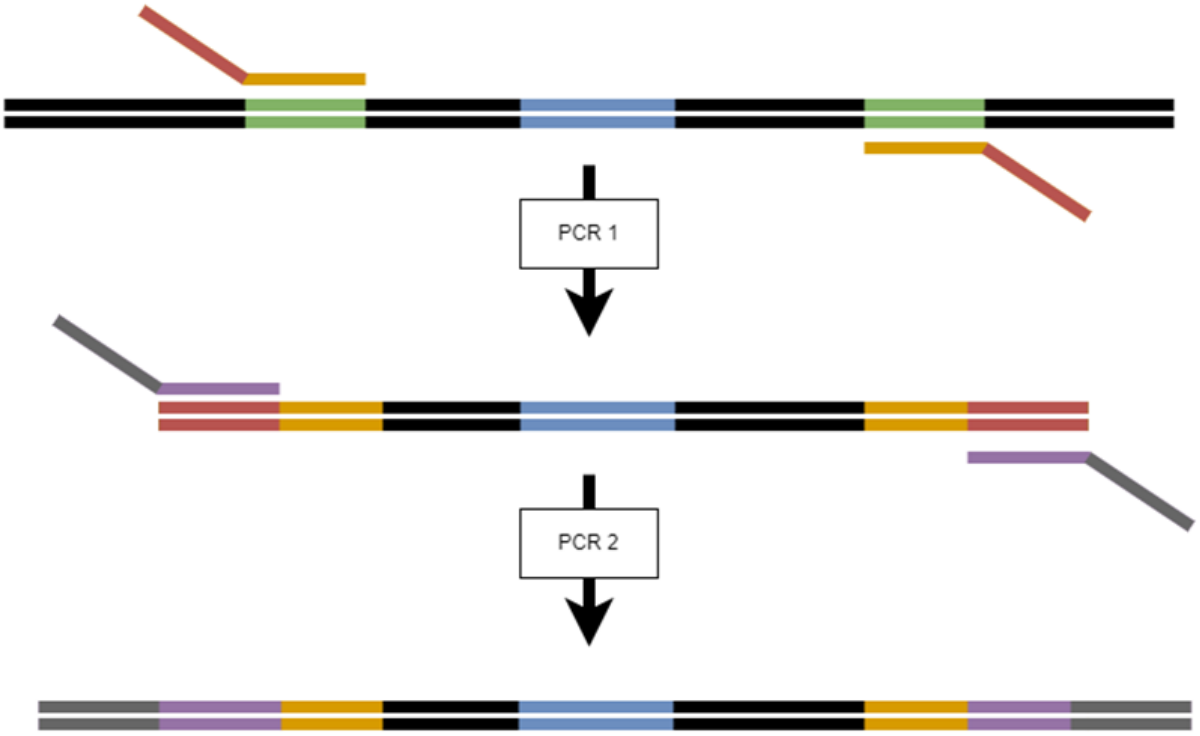
Both fragment analysis and amplicon sequencing approaches were unable to accurately determine variability within either the SSR1 or SSR2 homopolymer sequences. Both methods identified repeat sizes that did not agree with the expected standards set by previous WGS work and are not recommended for further use in MAP SSR calling or MSI identification. Further development of the amplicon sequencing method may be performed through the use of a higher resolution sequencing method such as NovaSeq, which may reduce the observed error rates shown for MiSeq-based methods.

Appendix 3.5: Figures and Tables for Appendix III

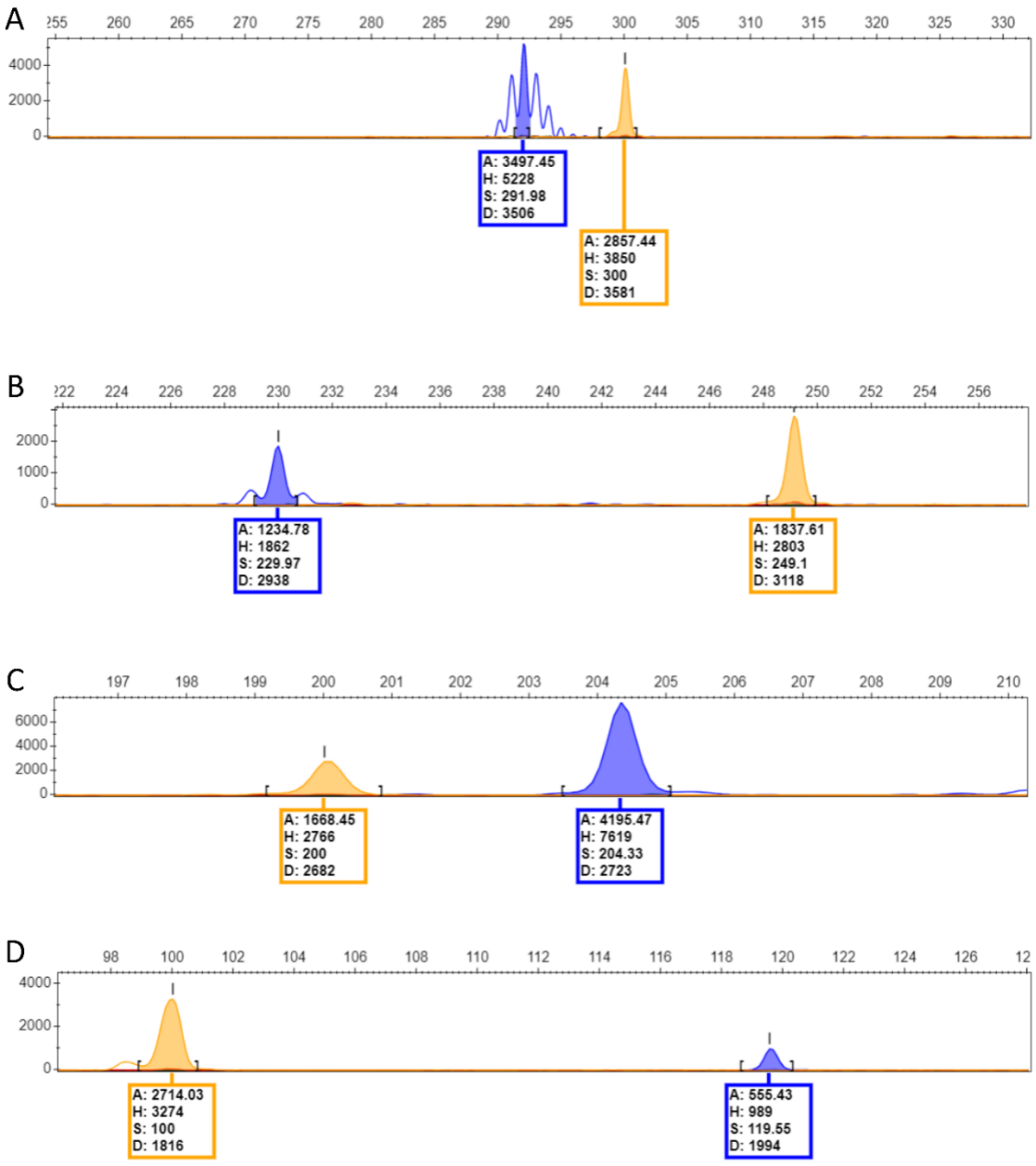
Appendix 3.5.1: Figures for Appendix III



Appendix Figure A5: DNA extraction protocol for use in fragment analysis assays. Fecal DNA was extracted directly from a portion of the unprocessed MAP-positive fecal sample and contained DNA from both MAP and non-MAP sources. After decontamination of the fecal sample, inoculation on MAP selective media, and growth of MAP colonies on solid media, isolates were either selected for growth in axenic cultures or collectively underwent DNA extraction (lawn DNA). Each isolate grown in axenic culture in liquid media underwent individual DNA extraction (axenic DNA).



Appendix Figure A6: PCR approach for amplicon sequencing. The repetitive element in each SSR locus (blue) is first amplified by targeting primer binding sites 5' and 3' to the repeat (green). The first PCR uses primers that target the primer binding sites (orange) while also containing a 3' tail which does not target the target sequence (red) **(Appendix Table A15)**. The second PCR uses primers that have a complementary sequence to the 3' tail of the PCR primers (purple) and have unique indexing sequences for identification in next-generation sequencing runs (grey) **(Appendix Table A16)**.



Appendix Figure A7: Electropherograms showing the fragment analysis results of SSR loci 1 (A), 2 (B), 8 (C) and 9 (D) using MAP strain K-10. Yellow peaks represent standard peaks produced by the GeneScan™ 500 LIZ™ dye size standard, while blue peaks represent amplicons produced using the 6-FAM labelled primer set. Peaks are described by the area under the peak (A), the height of the peak (H), the calculated size of the fragment (S), and the data point used to calculate each peak size (D). While SSRs 2, 8 and 9 show single peaks for each result, SSR1 shows additional shoulder peaks, suggesting additional fragments.

Appendix 3.5.2: Tables for Appendix III

Appendix Table A13: PCR protocols for amplification of tested SSR loci for the fragment analysis assay.

A: PCR Protocol for SSR1 Amplification

PCR Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98°C	30 seconds	1 cycle
Denaturation	98°C	10 seconds	
Annealing	64°C	30 seconds	40 cycles
Extension	72°C	30 seconds	
Final Extension	72°C	6 minutes	1 cycle

B: PCR Protocol for SSR2 Amplification

PCR Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98°C	30 seconds	1 cycle
Denaturation	98°C	10 seconds	
Annealing	55°C	30 seconds	10 cycles
Extension	72°C	30 seconds	
Denaturation	98°C	10 seconds	
Annealing	62°C	30 seconds	40 cycles
Extension	72°C	30 seconds	
Final Extension	72°C	3 minutes	1 cycle

C: PCR Protocol for SSR8 Amplification

PCR Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98°C	30 seconds	1 cycle
Denaturation	98°C	10 seconds	
Annealing	55°C	30 seconds	10 cycles
Extension	72°C	30 seconds	
Denaturation	98°C	10 seconds	
Annealing	61°C	30 seconds	40 cycles
Extension	72°C	30 seconds	
Final Extension	72°C	3 minutes	1 cycle

D: PCR Protocol for SSR9 Amplification

PCR Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98°C	30 seconds	1 cycle
Denaturation	98°C	10 seconds	
Annealing	65°C	30 seconds	40 cycles
Extension	72°C	30 seconds	
Final Extension	72°C	4 minutes 20 seconds	1 cycle

Appendix Table A14: Fragment analysis primer statistics as calculated by OligoCalc and mapping on the K-10 Genome.

Primer Name[a]	Primer Sequence (5'-3')	Nearest Neighbor Tm (°C)	CG (%)	# of nucleotides	#A	#T	#C	#G	Binding location (K-10 genome)	Amplicon Size (K-10)	Source
SSR2F*	TCGCCTC AGGCTTT ACTGAT	54.28	50	20	3	7	6	4	2,718,981 - 2,719,000	235 bp	(Podder et al. 2015)
SSR2R	CACGTAG GTCCGCT GATGA	53.83	58	19	4	4	5	6	2,719,179 - 2,719,215		(Podder et al. 2015)
SSR8F*	AGGCCTT CTACGTG CACAAC	54.3	55	20	5	4	7	4	1,028,049 - 1,028,068	208 bp	(Podder et al. 2015)
SSR8R	GAGATGT CCAGCCC TGTCTC	53.63	60	20	3	5	7	5	1,028,237 - 1,028,256		(Podder et al. 2015)
SSR9F*	CTCGTGG AAACCCT CGAC	51.12	61	18	4	3	7	4	2,955,316 - 2,955,333	127 bp	(Podder et al. 2015)
SSR9R	GGTGCTG AAATCCG GTGT	50.07	56	18	3	5	3	7	2,955,425 - 2,955,442		(Podder et al. 2015)
MAP_SSR 1_P5F*	CAGCTGA TAGTCGT TGGGCA	55.18	55	20	4	5	4	7	1,792,895 - 1,792,914	304 bp	None, Primer Blast Developed
MAP_SSR 1_P5R	GTACACC TGCAAGG GCGAG	54.58	63	19	5	2	5	7	1,793,180 - 1,793,198		None, Primer Blast Developed

[a]*All primers marked with "*" are attached with 6-FAM at the 5' end of the sequence when used for fragment analysis.*

Appendix Table A15: Primers used for the Amplification step of Amplicon Sequencing

Primer Name	Primer Sequence[a]	Nearest Neighbor Tm (°C)[b]	CG (%)[b]	# of nucleotides[b]	#A[b]	#T[b]	#C[b]	#G[b]	Binding location (K-10)[c]	Amplicon Size (K-10)[b]
SSR1c-III-F	ACACTCTTTCCCTACACGACGC TCTTCCGATCT**AGTGCACGTAC GGCCCCCCC**	59.47 (75.98)	75 (60)	20 (53)	3 (9)	2 (12)	10 (24)	5 (8)	1,793,077 - 1,793,096	150 bp (217 bp)
SSR1c-III-R	GTGACTGGAGTTCAGACGTGTG CTCTTCCGATCT**TCACCGACGG CGCCCCGATC**	61.61 (77.05)	75 (61)	20 (54)	3 (8)	2 (13)	10 (18)	5 (15)	1,793,207 - 1,793,226	
SSR1d-III-F	ACACTCTTTCCCTACACGACGC TCTTCCGATCT**ACGTGCTGCC GGAGCC**	55.04 (76.1)	75 (59)	16 (49)	2 (8)	2 (12)	6 (20)	6 (9)	1,793,038 - 1,793,053	121 bp (188 bp)
SSR1d-III-R	GTGACTGGAGTTCAGACGTGTG CTCTTCCGATCT**TGGTCCGCGC CAACGG**	54.98 (75.98)	75 (60)	16 (50)	2 (7)	2 (13)	6 (14)	6 (16)	1,793,143 - 1,793,158	

[a]*Portion of sequenced underlined and in bold represents the portion of the primer binding to the genome.*

[b]*Primer Binding Portion (Total Primer).*

[c]*Based on the primer binding portion of primer on the K-10 genome.*

Appendix Table A16: Indexing primers used for amplicon sequencing

Primer Name	Primer Sequence	Nearest Neighbor Tm (°C)	CG (%)	# of nucleotides	#A	#T	#C	#G
PCR-D711	CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGT	76.94	55	53	16	8	10	19
PCR-D712	CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGT	75.25	51	53	15	11	11	16
PCR-N726	CAAGCAGAAGACGGCATACGAGATGTCTTAGGGTGACTGGAGTTCAGACGTGT	74.22	51	53	15	11	9	18
PCR-N727	CAAGCAGAAGACGGCATACGAGATACTGATCGGTGACTGGAGTTCAGACGTGT	74.5	51	53	16	10	10	17
PCR-B707	CAAGCAGAAGACGGCATACGAGATAAGGCGTAGTGACTGGAGTTCAGACGTGT	74.78	51	53	17	9	9	18
PCR-B708	CAAGCAGAAGACGGCATACGAGATGTCCTAAGGTGACTGGAGTTCAGACGTGT	74.22	51	53	16	10	10	17
PCR-B723	CAAGCAGAAGACGGCATACGAGATCCGTTATGGTGACTGGAGTTCAGACGTGT	74.22	51	53	15	11	10	17
PCR-B724	CAAGCAGAAGACGGCATACGAGATTGCAAGACGTGACTGGAGTTCAGACGTGT	75.07	51	53	17	9	10	17
PCR-D507	AATGATACGGCGACCACCGAGATCTACACCAGGACGTACACTCTTTCCCTACACGAC	74.11	53	57	17	10	20	10
PCR-D508	AATGATACGGCGACCACCGAGATCTACACGTACTGACACACTCTTTCCCTACACGAC	73.81	51	57	17	11	20	9
PCR-S521	AATGATACGGCGACCACCGAGATCTACACGAGCCTTAACACTCTTTCCCTACACGAC	74.33	51	57	17	11	20	9
PCR-S522	AATGATACGGCGACCACCGAGATCTACACTTATGCGAACACTCTTTCCCTACACGAC	73.92	49	57	17	12	19	9
PCR-B501	AATGATACGGCGACCACCGAGATCTACACTTGAGCTCACACTCTTTCCCTACACGAC	74.58	51	57	16	12	20	9
PCR-B502	AATGATACGGCGACCACCGAGATCTACACACCGCTATACACTCTTTCCCTACACGAC	74.12	51	57	17	11	21	8
PCR-B515	AATGATACGGCGACCACCGAGATCTACACGCAACCATACACTCTTTCCCTACACGAC	74.13	51	57	18	10	21	8
PCR-B516	AATGATACGGCGACCACCGAGATCTACACAGTTGTGCACACTCTTTCCCTACACGAC	74.22	51	57	16	12	19	10

Appendix Table A17: Differences observed in the expected and observed sizes for PCR products targeting select SSR loci in MAP reference strain K-10.

SSR Locus	Repeat type	Number of Repeats	Expected Fragment Analysis Product Size (bp)	Observed Fragment Analysis Size (bp)
1	Mononucleotide (G)	19	304	292
2	Mononucleotide (G)	10	235	230
8	Trinucleotide (GGT)	5	208	204
9	Trinucleotide (TGC)	5	127	119

Appendix Table A18: Fragment Analysis results of a selection of strains from 14 high-shedding dairy cattle (n=139) and a comparison to their expected sizes as identified by whole genome sequencing.

Animal ID	Isolate ID	SSR1 Peak Size based on accepted Fragment Analysis Results	SSR1 Repeat Size Based on comparison to K-10 Fragment Analysis Size	SSR1 Repeat Size based on WGS Sequencing	Number of Agreeing SSR1 Sizes	SSR2 Peak Size based on accepted Fragment Analysis Results	SSR2 Repeat Size Based on comparison to K-10 Fragment Analysis Size	SSR2 Repeat Size based on WGS Sequencing	Number of Agreeing SSR2 Sizes
Standard	K-10	292	19	19	1/1	230	10	10	1/1
A25	#23	290	17	12		230	10	10	
A25	#24	290	17	12		230	10	10	
A25	#25	294	21	16		229	9	9	
A25	#26	293	20	15		229	9	9	
A25	#27	290	17	12	0/10	230	10	10	10/10
A25	#28b	291	18	13		230	10	10	
A25	#29b	294	21	16		232	12	12	
A25	#30b	289	16	11		230	10	10	
A25	#212	296	23	19		229	9	9	
A25	#213	296	23	20		230	10	10	
A26	#32	289	16	11		230	10	10	
A26	#33	291	18	13		229	9	9	
A26	#34	290/291	17/18	13		232	12	12	
A26	#35	289	16	11		231	11	11	
A26	#36	288/289	15/16	11	0/10	231	11	11	10/10
A26	#37	293	20	15		231	11	11	
A26	#38	290	17	12		230	10	10	
A26	#214	288	15	10		230	10	10	
A26	#215	289/290	16/17	12		230	10	10	

A26	#216	294	21	16		230	10	10	
AM1	#217	296	23	18		231	11	10	
AM1	#218	292	19	14		231	11	10	
AM1	#219	289	16	11		231	11	10	
AM1	#220	289	16	11		232	12	11	
AM1	#221	295	22	18	0/10	231	11	10	2/10
AM1	#222	289/290	16/17	12		232	12	11	
AM1	#223	290	17	12		233	13	12	
AM1	#224	289	16	11		232	12	11	
AM1	#225	293	20	14		230	10	10	
AM1	#226	292/293	19/20	15		232	12	12	
A32	#113	295	22	18		233/234	13/14	14	
A32	#114	294/295	21/22	17		232	12	12	
A32	#115	292	19	14		236/237	16/17	UNK	
A32	#116	291	18	13		234	14	14	
A32	#117	293	20	15	0/10	233	13	13	5 or 6/10
A32	#118	294	21	17		234	14	15	
A32	#119b	296/297	23/24	18		233	13	16	
A32	#120	295	22	18		236	16	18	
A32	#121	292/293	19/20	14		232	12	12	
A32	#122	293	20	15		232	12	12	
AM2	#123	292	19	14		232	12	12	
AM2	#124	298	25	19		UNK	UNK	14	
AM2	#125	294/295	21/22	17		234/235	14/15	15	
AM2	#228	295/296	22/23	18	0/10	231	11	11	5 or 6/10
AM2	#229	294	21	15		234	14	14	
AM2	#230	299	26	20		233	13	14	
AM2	#231	296	23	12		233	13	10	
AM2	#232	297	24	20		232	12	12	

AM2	#233	295	22	19		233	13	13	
AM2	#234	295/296	22/23	18		231	11	13	
A34	#126	291	18	13		233	13	13	
A34	#127	298/299	25/26	19		231	11	11	
A34	#128	296	23	19		233	13	13	
A34	#129	295	22	17		230	10	10	
A34	#130	292	19	14	0/10	232	12	12	9/9 available results
A34	#131	292/293	19/20	14		UNK	UNK	14	
A34	#132	297	24	22		231	11	11	
A34	#133	293	20	15		231	11	11	
A34	#236	292	19	14		233	13	13	
A34	#237	291/292	18/19	14		233	13	13	
AM3	#134	298/299	25/26	20		234	14	15	
AM3	#135	298/299	25/26	17		232	12	12	
AM3	#238	298/299	25/26	UNK		234/235/236	14/15/16	14	
AM3	#239	295/296	22/23	19		230	10	10	
AM3	#240	296/297	23/24	20	0/9 available results	231	11	11	8 or 9/10
AM3	#241	294	21	16		231	11	11	
AM3	#242	297	24	22		233	13	13	
AM3	#243	294/295	21/22	17		233	13	13	
AM3	#244	298	25	22		234	14	14	
AM3	#245	297/298	24/25	22		233	13	13	
A36	#136	298/299	25/26	UNK		UNK	UNK	13	
A36	#247	295	22	12		230	10	10	
A36	#248	297/298	24/25	UNK		232	12	12	
A36	#249	298/299	25/26	24	0/5 available results	230/231	10/11	10	7 or 8/9 available results
A36	#250	297	24	UNK		231	11	11	
A36	#251	298/299	25/26	UNK		234	14	13	
A36	#252	295	22	19		230	10	10	

A36	#253	298	25	UNK		232	12	12	
A36	#254	294	21	20		230	10	10	
A36	#255	295	25	18		230	10	10	
A37	#137	298/299	25/26	20		230	10	10	
A37	#256	298/299	25/26	UNK		231	11	11	
A37	#258	298/299	25/26	UNK		230	10	10	
A37	#259	295/296	22/23	18		230	10	10	
A37	#260	298/299	25/26	UNK	0/7 available results	233	13	13	10/10
A37	#261	296/297	23/24	17		230	10	10	
A37	#262	292	19	13		230	10	10	
A37	#263	295/296	22/23	19		230	10	10	
A37	#264	293/294	20/21	16		230	10	10	
A37	#265	296/297	23/24	19		230	10	10	
A19	#159	294/295	21/22	18		230	10	10	
A19	#160	295	22	19		230	10	10	
A19	#161	293	20	15		229	9	9	
A19	#162	296	23	19		230	10	10	
A19	#163	UNK	UNK	UNK	0/9 available results	230	10	10	10/10
A19	#164	295	22	18		230	10	10	
A19	#165	290	17	12		233	13	13	
A19	#166	292	19	14		229	9	9	
A19	#167	295/296	22/23	18		230	10	10	
A19	#168	295/296	22/23	19		230	10	10	
A20	#169	295	22	17		232	12	12	
A20	#170	296/297	23/24	UNK		233/234	13/14	14	
A20	#171	294	21	17	0/8 available results	230	10	10	8 or 9/9
A20	#172	294	21	15		233	13	13	
A20	#173	293	20	15		233	13	13	
A20	#174	292	19	14		230	10	10	

A20	#175	290	17	12		230	10	10	
A20	#176	296/297	23/24	21		230	10	10	
A20	#177	290	17	12		232	12	12	
A21	#179	295	22	16		230	10	10	
A21	#180	297	24	20		230	10	10	
A21	#181	298/299	25/26	23		232	12	12	
A21	#182	294	21	15		234	14	14	
A21	#183	293	20	15	0/10	231	11	11	9/10
A21	#184	294	21	15		231	11	11	
A21	#185	296	23	18		231	11	11	
A21	#186	292/293	19/20	17		231	11	12	
A21	#187	293	20	15		231	11	11	
A21	#188	295/296	22/23	19		232	12	12	
A22	#189	293	20	14		234	14	14	
A22	#190	294	21	15		231	11	11	
A22	#191	293	20	14		230	10	10	
A22	#192	294/295	21/22	17		231	11	11	
A22	#193	293	20	14	0/10	232	12	12	9/10
A22	#194	294/295	21/22	16		230	10	10	
A22	#195	293	20	14		232	12	12	
A22	#196	292/293	19/20	16		231	11	11	
A22	#197	296	23	19		234	14	13	
A22	#198	292	19	13		230	10	10	
A23	#199	294/295/296	21/22/23	15		236/237	16/17	15	
A23	#200	296	23	19		235/236	15/16	15	
A23	#201	290	17	12	0/10	237	17	18	4 to 7/10
A23	#202	293/294	20/21	15		231	11	11	
A23	#203	292	19	14		233	13	13	
A23	#204	293	20	15		236/237	16/17	16	

A23	#205	291	18	13	234	14	15
A23	#206	293	20	15	231	11	11
A23	#207	293	20	14	233/234	13/14	14
A23	#208	290	17	12	232	12	12

Appendix Table A19: Number of reads identified by amplicon sequencing of axenic cultures that matched expected repeat sizes previously identified by whole genome sequencing of isolates grown from high-shedding dairy cattle.

Animal ID	Repeat Size[a]	Number of Reads (Primer Set C)	Number of Reads (Primer Set D)
A23	7	14	2778
	8	0	2771
	9	119	2762
	10	864	2750
	11	1999	2617
	12	2798	2047
	13	1929	1062
	14	486	290
	15	41	29
	16	3	8
	17	0	2
	18	0	0
	19	0	0
	20	0	0
	21	0	0
A25	7	9	22
	8	20	24
	9	601	122
	10	3771	785
	11	3672	2357
	12	2458	2349
	13	1169	998
	14	293	268
	15	27	23
	16	3	1
	17	0	0
	18	0	0
	19	1	0
	20	0	0
	21	0	0
A26	7	7	14
	8	105	113
	9	1676	1685
	10	2272	2282
	11	3907	3918
	12	2529	2541

	13	1167	1180
	14	329	343
	15	23	38
	16	4	20
	17	0	17
	18	1	19
	19	0	19
	20	0	20
	21	0	21

[a]*Repeat sizes highlighted in blue were confirmed to be present in whole genome sequencing results of isolates that were pooled together for use in amplicon sequencing samples, while those highlighted in orange were not.*

Appendix Table A20: Number of reads identified by amplicon sequencing of lawn DNA grown from high-shedding dairy cattle.

Animal ID	Repeat Size	Number of Reads (Primer Set C)	Number of Reads (Primer Set D)
A23	7	3256	3263
	8	3249	3257
	9	3236	3245
	10	3223	3233
	11	3071	3082
	12	2352	2364
	13	1163	1176
	14	307	321
	15	40	55
	16	7	23
	17	1	18
	18	0	18
	19	0	19
	20	0	20
	21	0	21
A25	7	12	6
	8	6	69
	9	186	1158
	10	1523	1631
	11	4013	2590
	12	2917	1902
	13	1463	915
	14	304	218
	15	11	19
	16	2	4
	17	0	0
	18	0	0
	19	0	0
	20	0	0
	21	0	0
A26	7	9	13
	8	31	84
	9	531	1504
	10	3024	1939
	11	3623	3156
	12	2480	2372
	13	1118	1046

14	263	326
15	28	27
16	1	3
17	0	0
18	0	0
19	1	0
20	0	0
21	0	0

Bibliography for Appendix III

1. Podder, M. P., Banfield, S. E., Keefe, G. P., Whitney, H. G. & Tahlan, K. Typing of *Mycobacterium avium* subspecies *paratuberculosis* Isolates from Newfoundland Using Fragment Analysis. *PLOS ONE* **10**, e0126071 (2015).
2. Ahlstrom, C., Barkema, H. W. & De Buck, J. Improved Short-Sequence-Repeat Genotyping of *Mycobacterium avium* subsp. *paratuberculosis* by Using Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry. *Applied and Environmental Microbiology* **80**, 534–539 (2014).
3. Amonsin, A. *et al.* Multilocus Short Sequence Repeat Sequencing Approach for Differentiating among *Mycobacterium avium* subsp. *paratuberculosis* Strains. *J Clin Microbiol* **42**, 1694–1702 (2004).
4. Li, L. *et al.* The complete genome sequence of *Mycobacterium avium* subspecies paratuberculosis. *Proc Natl Acad Sci U S A* **102**, 12344–12349 (2005).
5. Wynne, J. W. *et al.* Resequencing the *Mycobacterium avium* subsp. *paratuberculosis* K10 genome: improved annotation and revised genome sequence. *J Bacteriol* **192**, 6319–6320 (2010).
6. Davidson, F. W., Ahlstrom, C., De Buck, J., Whitney, H. G. & Tahlan, K. Examination of *Mycobacterium avium* subspecies *paratuberculosis* mixed genotype infections in dairy animals using a whole genome sequencing approach. *PeerJ* **4**, e2793 (2016).
7. Oakey, J., Gavey, L., Singh, S. V., Platell, J. & Waltisbuhl, D. Variable-number tandem repeats genotyping used to aid and inform management strategies for a bovine

Johne's disease incursion in tropical and subtropical Australia. *J Vet Diagn Invest* **26**, 651–657 (2014).

8. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115 (2012).
9. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
10. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35**, W43-46 (2007).
11. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671–675 (2012).
12. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
13. Life Technologies. DNA Fragment Analysis by Capillary Electrophoresis. (2014).
14. Fazekas, A., Steeves, R. & Newmaster, S. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* **48**, 277–285 (2010).
15. Kieleczawa, J. Fundamentals of Sequencing of Difficult Templates—An Overview. *J Biomol Tech* **17**, 207–217 (2006).
16. Sohal, J. S. *et al.* Genetic structure of *Mycobacterium avium* subsp. *paratuberculosis* population in cattle herds in Quebec as revealed by using a combination of multilocus genomic analyses. *J Clin Microbiol* **52**, 2764–2775 (2014).
17. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics and Bioinformatics* **3**, lqab019 (2021).

Appendix IV: Data and Code Availability

Code used for each Chapter of this thesis is available at the following repository:

https://gitfront.io/r/asb068/amhz1zp7hQfQ/Alex-Byrne-Thesis-Repository/

The 67 datasets presented in Chapter IV of this thesis are associated with the NCBI BioProject PRJNA925907 and can be found in the NCBI Sequence Read Archive (SRA) repository under the accession numbers SRR23179790 to SRR23179856.

The 139 MAP genome sequences presented in Chapter V of this thesis are associated with BioProject PRJNA925907 and have been deposited in the NCBI SRA repository under the accession numbers: SRR23179790-SRR23179792, SRR23179800, SRR23179819, SRR23179829, SRR23179835, SRR23179843, SRR23179850, SRR23179851, SRR23179853, SRR23179854, SRR23179855, SRR23179856, and SRR24326188-SRR24326312. References and accession numbers for all other publicly available datasets used are presented in **Appendix Table A1**.

Bibliography for Appendix IV

1. Alonge, M. et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology 23, 258 (2022).

2. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455–477 (2012).

3. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890 (2018).

4. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Research 43, e15 (2015).

5. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075 (2013).

6. Lu, J. KrakenTools. (2020).

7. Minh, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37, 1530–1534 (2020).

8. National Center for Biotechnology Information. SRA Toolkit. (2022).

9. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution 32, 268–274 (2015).

10. Ozer, E. A. IN_SILICO_PCR. (2022).

11. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2, e000056 (2016).

12. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015).

13. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70, e102 (2020).

14. Sailer, Z. & Silvester, S. Jupyter Notebook. (2022).

15. Seemann, T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014).

16. Seemann, T. Snippy: fast bacterial variant calling from NGS reads. (2015).

17. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol 20, 257 (2019).