



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome-wide diversity and phylogeography of *Mycobacterium avium* subsp. *paratuberculosis* in Canadian dairy cattle

Citation for published version:

Ahlstrom, C, Barkema, HW, Stevenson, K, Zadoks, RN, Biek, R, Kao, R, Trewby, H, Hauptstein, D, Kelton, DF, Fecteau, G, Labrecque, O, Keefe, GP, McKenna, SLB, Tahlan, K & De Buck, J 2016, 'Genome-wide diversity and phylogeography of *Mycobacterium avium* subsp. *paratuberculosis* in Canadian dairy cattle', *PLoS ONE*, vol. 11, no. 2, e0149017. <https://doi.org/10.1371/journal.pone.0149017>

Digital Object Identifier (DOI):

[10.1371/journal.pone.0149017](https://doi.org/10.1371/journal.pone.0149017)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

PLoS ONE

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



RESEARCH ARTICLE

Genome-Wide Diversity and Phylogeography of *Mycobacterium avium* subsp. *paratuberculosis* in Canadian Dairy Cattle

Christina Ahlstrom¹, Herman W. Barkema¹, Karen Stevenson², Ruth N. Zadoks^{2,3}, Roman Biek³, Rowland Kao³, Hannah Trewby³, Deb Hauptstein⁴, David F. Kelton⁵, Gilles Fecteau⁶, Olivia Labrecque⁷, Greg P. Keefe⁸, Shawn L. B. McKenna⁸, Kapil Tahlan⁹, Jeroen De Buck^{1*}

1 University of Calgary, Calgary, Alberta, Canada, **2** Moredun Research Institute, Penicuik, Scotland, **3** University of Glasgow, Glasgow, Scotland, **4** SaskMilk, Regina, Saskatchewan, Canada, **5** University of Guelph, Guelph, Ontario, Canada, **6** Université de Montréal, Montréal, Québec, Canada, **7** Laboratoire d'épidémiologie animale du Québec, Saint-Hyacinthe, Québec, Canada, **8** University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada, **9** Memorial University of Newfoundland and Labrador, St. John's, Newfoundland, Canada

* jdebuck@ucalgary.ca



OPEN ACCESS

Citation: Ahlstrom C, Barkema HW, Stevenson K, Zadoks RN, Biek R, Kao R, et al. (2016) Genome-Wide Diversity and Phylogeography of *Mycobacterium avium* subsp. *paratuberculosis* in Canadian Dairy Cattle. PLoS ONE 11(2): e0149017. doi:10.1371/journal.pone.0149017

Editor: Srinand Sreevatsan, University of Minnesota, UNITED STATES

Received: October 20, 2015

Accepted: January 25, 2016

Published: February 12, 2016

Copyright: © 2016 Ahlstrom et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Files are available from the NCBI Bioproject, accession number PRJNA274491 (<http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA274491>).

Funding: This research was supported by the Alberta Livestock and Meat Agency and Dairy Farmers of Canada. SaskMilk provided support in the form of salaries for author DH, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of all authors are articulated in the 'author contributions' section.

Abstract

Mycobacterium avium subsp. *paratuberculosis* (MAP) is the causative bacterium of Johne's disease (JD) in ruminants. The control of JD in the dairy industry is challenging, but can be improved with a better understanding of the diversity and distribution of MAP subtypes. Previously established molecular typing techniques used to differentiate MAP have not been sufficiently discriminatory and/or reliable to accurately assess the population structure. In this study, the genetic diversity of 182 MAP isolates representing all Canadian provinces was compared to the known global diversity, using single nucleotide polymorphisms identified through whole genome sequencing. MAP isolates from Canada represented a subset of the known global diversity, as there were global isolates intermingled with Canadian isolates, as well as multiple global subtypes that were not found in Canada. One Type III and six "Bison type" isolates were found in Canada as well as one Type II subtype that represented 86% of all Canadian isolates. Rarefaction estimated larger subtype richness in Québec than in other Canadian provinces using a strict definition of MAP subtypes and lower subtype richness in the Atlantic region using a relaxed definition. Significant phylogeographic clustering was observed at the inter-provincial but not at the intra-provincial level, although most major clades were found in all provinces. The large number of shared subtypes among provinces suggests that cattle movement is a major driver of MAP transmission at the herd level, which is further supported by the lack of spatial clustering on an intra-provincial scale.

Competing Interests: DH is employed by SaskMilk, a non-profit organization representing Saskatchewan dairy producers. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

Introduction

Mycobacterium avium subsp. *paratuberculosis* (MAP) is the causative bacterium of Johne's disease (JD) in ruminants and is widespread in the global dairy industry [1]. A potential association between MAP and Crohn's disease in humans is increasing pressure to reduce the prevalence in dairy herds by breaking the transmission cycle [2,3]. This can be achieved within herds by eliminating exposure of calves to infected feed and feces and between herds by preventing the introduction of MAP-infected animals [4]. Despite this knowledge, the herd-level prevalence of JD in Canada remains high [5].

MAP is a Gram-positive, acid-fast bacillus classified into two broad strain types—Type I and Type II [6]. Additional types have also been described, however Type III is now understood to be a sub-lineage of Type I [7] and “Bison type” (Type B) is a sub-lineage of Type II (Bryant et al., submitted for publication). Subtyping of isolates beyond the broad strain types has largely relied on repetitive regions of the genome; however, this limits the ability to infer true evolutionary relationships ([8]; Bryant et al., submitted for publication). The relatively high discriminatory ability of short sequence repeat (SSR) [9] and variable number tandem repeat (VNTR) typing [10] has motivated researchers to use these tools to investigate the diversity of MAP in several countries and host species (see [11] for a review). In Canada, the genetic structure of MAP was analyzed at a provincial level using a combination of SSR and VNTR loci [12]. Two genotypes accounted for 43% of the isolates and the majority of the remaining isolates differed by a single, potentially unstable, locus. Thus, more appropriate genotyping tools are required to capture the true population diversity necessary for molecular epidemiological analyses. Whole genome sequencing (WGS) provides a means to identify informative, stable molecular markers that can subsequently be used to assess the population structure in a broader scope [13].

A thorough understanding of the population structure of mycobacteria at multiple geographical scales can provide an important framework to investigate transmission in an epidemiological context as well as subtype-specific characteristics, such as virulence, transmissibility, and immunogenicity. WGS has been used to study the persistence and recent transmission events of *M. bovis* in cattle and badgers in UK dairy herds, improving the understanding of disease dynamics at small spatial scales [14]. Disease diversity has also been attributed to strain diversity in a number of pathogens, including *Mycobacterium tuberculosis* [15]. Strain-specific differences in virulence and pathogenicity of MAP, however, have mainly focused on the major strain types [16–19], in which significant differences were found between Type I and Type II isolates in their growth rates and intracellular survival.

An understanding of the true genetic diversity of MAP is needed to accurately assess subtype-specific phenotypes and virulence characteristics. Additionally, rational vaccine development and potentially effective management practices rely on a comprehensive assessment of circulating MAP subtypes and their distribution across different provinces. The objectives of this study were to A) determine the diversity and evolutionary relationship of MAP strain types in Canadian dairy farms from all 10 provinces using WGS, B) contextualize this diversity through comparison to a diverse global collection of MAP isolates, and C) assess the phylogeography of Canadian MAP isolates on a national and provincial level.

Materials and Methods

Canadian isolate selection and DNA preparation

Canadian MAP isolates were obtained from environmental and individual cow fecal samples collected from regional JD control initiatives in British Columbia, Alberta, Saskatchewan,

Ontario, Québec, and the Atlantic region (New Brunswick, Nova Scotia, Prince Edward Island and Newfoundland), as detailed previously [5,8,20]. Additional MAP isolates were obtained from the Alberta Ministry of Agriculture and Rural Development ($n = 22$), representing individual cow fecal samples from different dairy herds in Alberta sampled between 1999 and 2002. Environmental manure samples from manure storage areas were also obtained from five dairy operations in the province of Manitoba. MAP DNA from environmental manure samples from three dairy herds in Newfoundland was also obtained. The isolates included in this study were obtained from a total of 170 dairy herds, all with permission from the owners, and are presented in [S1 Table](#). Only one isolate per herd was selected, with the exception of 11 herds (located in Alberta or Saskatchewan) in which two or three isolates were included that previous WGS analyses identified as genetically distinct [8]. MAP was grown from environmental manure samples and individual cow fecal samples using the TREK ESP Culture System reagents (TREK Diagnostics, Cleveland, OH, USA) in all provinces, with the exception of Québec, where isolates were cultured with the BACTEC MGIT 960 ParaTB culture system (Becton, Dickinson and company, Franklin Lakes, NJ, USA). Liquid culture broth of IS900 PCR positive samples [5,21] was plated onto Middlebrook 7H11 agar supplemented with OADC and 2 mg/L mycobactin J and a single MAP colony was subcultured for subsequent regrowth and genomic DNA extraction, as detailed previously [22]. Multiplexed DNA libraries were prepared using the Nextera XT sample preparation kit (Illumina, San Diego, CA, USA) according to manufacturer's instructions.

Whole genome sequencing and single nucleotide polymorphism (SNP) detection

Paired-end WGS was performed on pools of 24 samples using the MiSeq platform (Illumina, San Diego, CA, USA) with a read length of either 250 or 300 base pairs. Reads were deposited in the sequence read archive (accession number SRP067719 and SRR060191). BWA [23] was used for reference mapping to the revised version of the MAP K10 genome (NCBI Sequence Read Archive study SRR060191) [24,25] and variant sites were identified using SAMtools and bcftools [26]. Variant sites within insertion elements and PE/PPE regions were discounted and subsequent filtering criteria [8] were used to minimize false-positives. Filtering criteria ensured at least 2 high quality SNPs on both the forward and reverse strands, a mapping quality >40 , and a heterozygosity $<5\%$ to assign a position as either the "reference" or "alternate" allele. Isolates that failed the filtering criteria at a particular site was treated as an "unknown base" and assigned a value of "N".

Global sequence selection

MAP isolates ($n = 135$) from 22 countries were collected at Moredun Research Institute (MRI) having been selected in order to maximize the global genetic diversity. WGS and data analysis was performed as detailed elsewhere (Bryant et al., submitted for publication). A total of 26 sequences were selected from this collection for comparison to sequences from Canadian isolates based on their phylogenetic position and availability of metadata, such as year of isolation and host. All major phylogenetic clades were represented, including MAP types I, II, III, and B and raw sequencing reads were processed alongside the Canadian sequences, as described above. Additional sequences from the MRI global collection were subsequently analyzed using the same methods in a separate analysis to refine Type III and Type B isolates in the phylogeny.

Phylogenetic analyses

Concatenated variant sites were used for all phylogenetic analyses. Maximum likelihood phylogenetic trees were created using PhyML [27] with the TPM1uf nucleotide substitution model,

as determined by jModelTest (version 2.1.7) [28]. Node support was evaluated with 100 bootstrap pseudoreplications. Phylogenetic trees were visualized using FigTree (version 1.4.2), and iTOL [29] was used to annotate the tips with geographical location.

The phylogenetic signal was evaluated using the likelihood mapping function in TREE-PUZZLE [30,31]. This method determines the support for internal branches of a phylogeny and the phylogenetic content of a dataset by plotting 10,000 randomly selected quartets in a likelihood-mapping diagram. The percentage of quartets that cannot be reliably resolved demonstrates the strength of the phylogenetic signal. Additionally, the nucleotide alignment was tested for the presence of recombinant sequences using the pairwise homoplasy index (PHI) test [32] in SplitsTree (version 4.13.1) [33] with a significance threshold of $p \leq 0.05$. A rarefaction analysis was performed to determine the subtype richness among different Canadian provinces. Subtypes, or clades, were assigned using a strict and relaxed definition; the strict definition divided the isolates into nine divergent subtypes based on a shared branch length of at least 50 SNPs and the relaxed definition, selected to capture the majority of subpopulations in the phylogeny, divided the isolates into 32 subtypes. The function RAREFY in the package VEGAN (version 2.2–1) was implemented in R, version 3.1.2 [34].

The phylogenetic association between MAP and geographical location was tested using BaTS (Bayesian Tip Significance Testing) [35]. This program tests the null hypothesis of no correlation between phylogeny and geographic location by performing randomization tests from a posterior set of trees generated through a Bayesian Markov chain Monte Carlo analysis. A posterior set of trees was generated using the general time-reversible (GTR) substitution model, a lognormal relaxed molecular clock, and a constant-sized coalescent model implemented in BEAST (version 1.8.0) [36]. BaTS takes into consideration phylogenetic uncertainty when testing the association index, parsimony score, and the monophyletic clade size statistic. The phylogenetic association was tested with province of origin as well as with three regions (each between 43,000 and 45,000 square Km) within Alberta for which the location of farm origin was known ($n = 58$; Region 1, north of Red Deer; Region 2, between Red Deer and Calgary; Region 3, south of Calgary). A p -value ≤ 0.05 was regarded as statistically significant. Exact binomial tests were performed to test the underrepresentation of regions in particular phylogenetic clades using the exact binomial test implemented in R [34].

Results

Canadian and global isolates

A total of 9,670 variant sites were identified in 182 MAP isolates from Canada and 26 global isolates. The average depth of coverage of the Canadian and global isolates was 29 and 139, respectively; the average depth of coverage and number of raw reads for each isolate is presented in [S1 Table](#). Results from the maximum likelihood phylogenetic analysis of concatenated variant sites are presented in [Fig 1](#). One isolate from Québec (A1_536_QC) was most closely related to a Type III isolate from the global collection (MAPMRI051), with 1,411 pairwise SNP differences. The Canadian Type III sequence was subsequently compared to an additional nine Type III sequences from the global collection, which identified a deer isolate from the Czech Republic (A1_536_QC) as the closest relative (313 pairwise SNP differences) ([Fig 2A](#)).

The remaining 181 Canadian isolates clustered with the Type II global isolates. A dominant subtype included 155 of the Canadian and five of the global isolates. Additionally, six Canadian isolates from three provinces (Alberta, Ontario, and Québec) were most closely related to the Type B isolate from the global collection (MAPMRI0128). To further investigate the relationship between all Type B isolates, an additional three global Type B sequences (from isolates

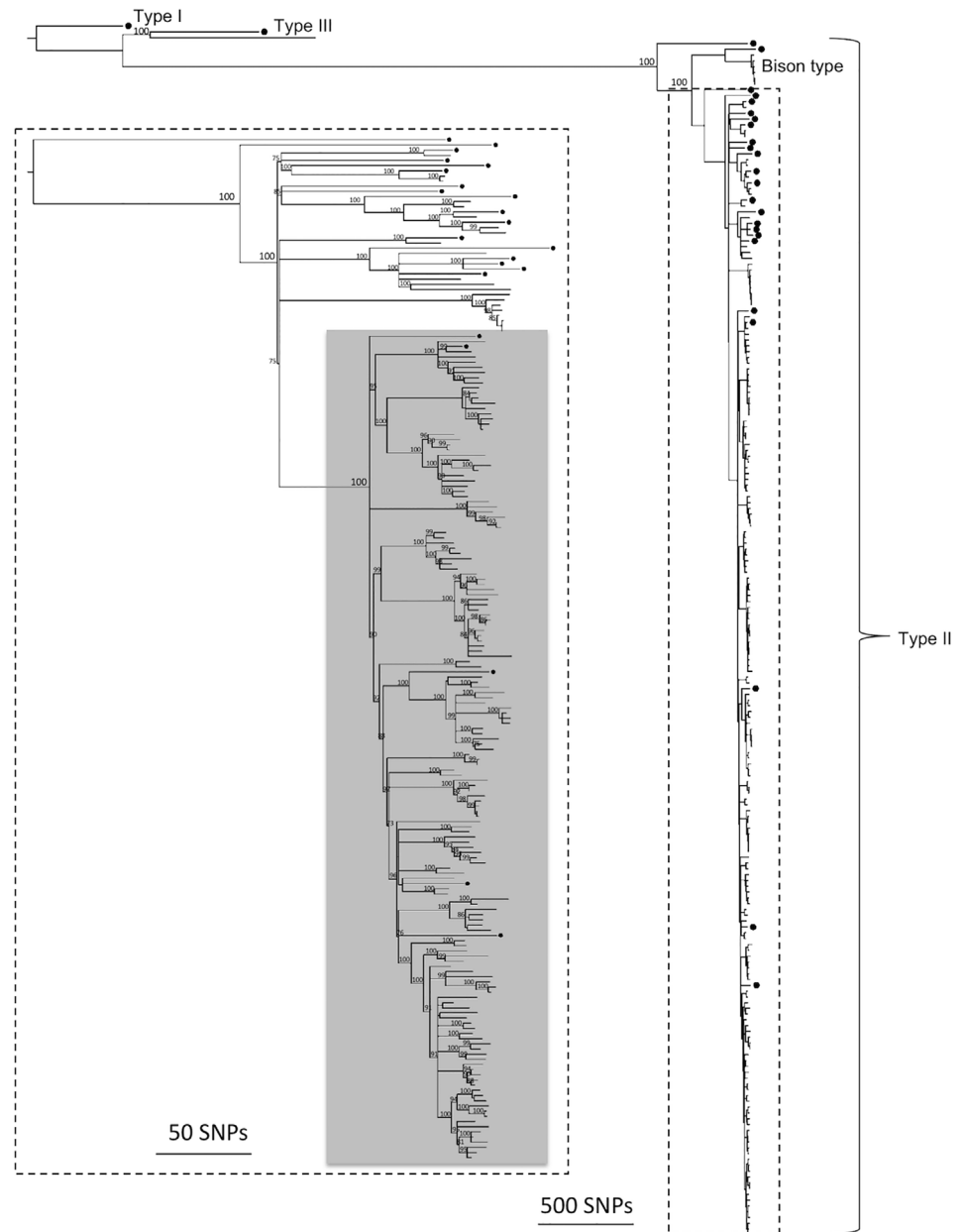


Fig 1. Maximum likelihood phylogenetic tree of sequenced MAP isolates. Phylogenetic tree of global ($n = 26$, labeled with a black dot (•)) and Canadian ($n = 182$) isolates based on 9,670 variant sites using the TPM1 of nucleotide substitution model. The tree is rooted to the Type I sequence. A magnification of the phylogeny excluding the Type I, III, and B isolates is outlined in dotted lines. The dominant subtype is shaded in grey. Bootstrap values with branch support $\geq 70\%$ are displayed.

doi:10.1371/journal.pone.0149017.g001

originating from India, the Czech Republic, and the USA) were analyzed with the six Canadian sequences. A total of 403 SNPs were identified among Type B sequences and the phylogenetic analysis based on these SNPs revealed close similarity between the Canadian isolates and an isolate originating from a bison herd in the USA (MAPMRI029) (Fig 2B). The USA and Canadian sequences had between eight and 31 pairwise SNP differences, with the USA sequence most closely related to the sequence from Alberta (A1_028_AB).

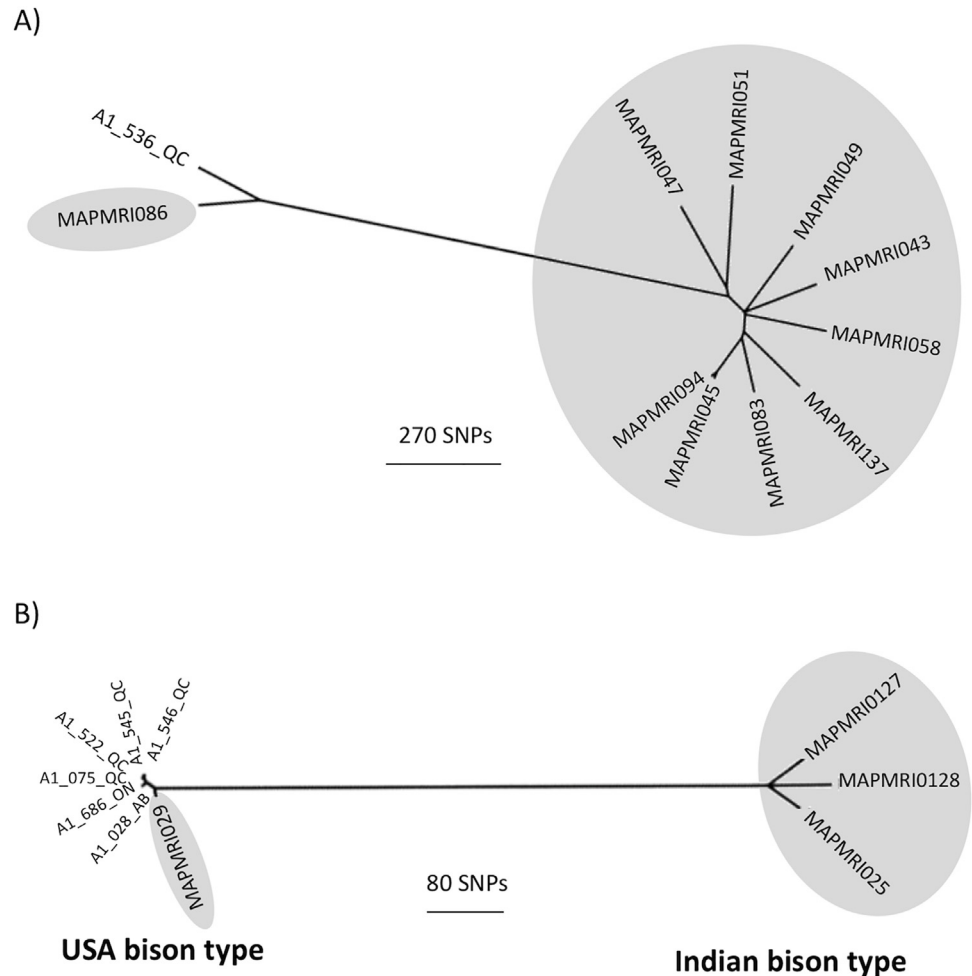


Fig 2. Maximum likelihood phylogenetic trees of Canadian and global MAP isolates. A) Phylogenetic tree of one Canadian and ten global Type III isolates based on 5,416 concatenated variant sites. B) Phylogenetic tree of six Canadian and four global "Bison type" isolates based on 403 concatenated variant sites. The global sequences are shaded in grey and branches are labeled according to the isolate ID.

doi:10.1371/journal.pone.0149017.g002

Canadian isolates

6,604 variant sites were identified in 182 Canadian isolates, although nearly half (3,022) were unique to the Type III isolate. 3,582 SNPs were identified within the Type II isolates, including the six Type B isolates. Likelihood mapping analysis indicated a strong phylogenetic signal, with greater than 80% of the quartets fully resolved (S1 Fig). The PHI test did not indicate statistically significant evidence for recombination, with $p = 0.73$.

A total of nine divergent MAP subtypes were found in Canada, of which eight were classified as Type II. Within the Type II isolates, 86% belonged to the dominant subtype (subtype H), with the remaining subtypes each representing less than 4% of the population (subtype A, B, C, D, E, F, and G represented 3, 0.5, 1, 3, 2, 4, 0.5% respectively). Two of those subtypes (B and G) were each represented by a single isolate. Subtypes that contained more than one isolate included as few as one region or as many as all seven. Within the dominant subtype in particular, long internal branches differentiated this subtype into several subpopulations that were differentiated in the rarefaction analyses using the relaxed subtype definition.

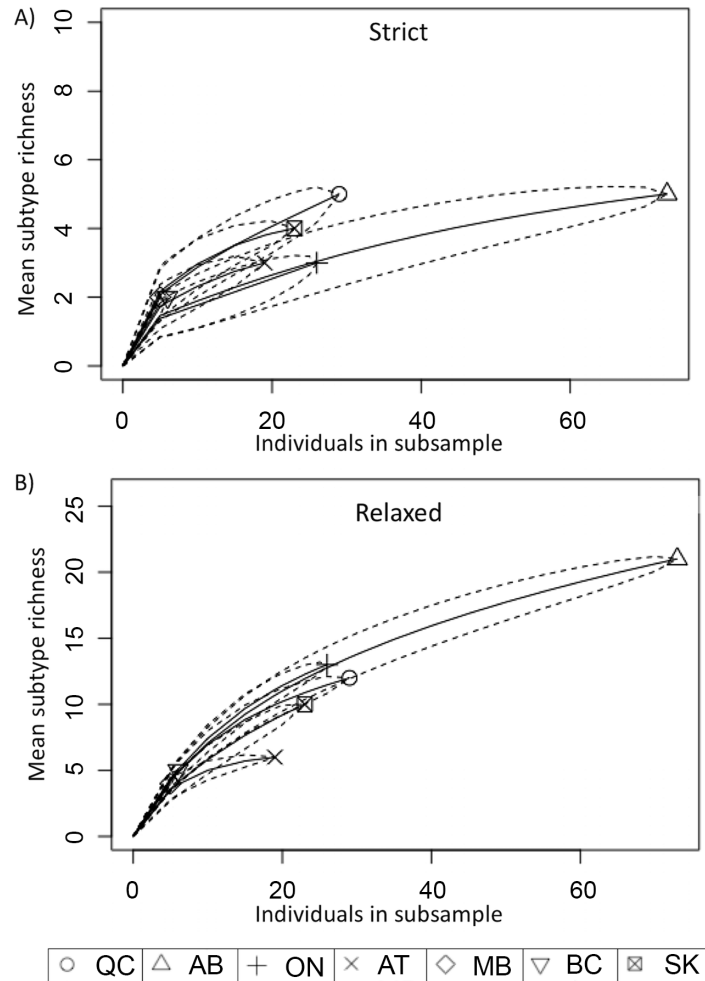


Fig 3. Rarefaction curves indicating the mean subtype richness of each region at different sampling efforts. Rarefaction curves were generated using the function rarefy in the R package Vegan using A) a strict subtype definition (9 total subtypes) and B) a relaxed subtype definition (32 total subtypes). 95% confidence intervals are indicated by a dotted line. The 7 regions are labeled with shapes indicated in the key at the bottom of the figure. Regions are abbreviated as follows: Québec = QC, Alberta = AB, Ontario = ON, the Atlantic region = AT, Manitoba = MB, British Columbia = BC, and Saskatchewan = SK.

doi:10.1371/journal.pone.0149017.g003

The rarefaction analyses quantified the subtype richness at two different cutoffs to capture the diversity based on more recent and more ancestral SNPs given a range of sampling efforts for each region (Fig 3). Using the strict, more ancestral subtype definition Québec displayed more diversity than Alberta at a sampling effort of 25 samples. However, at smaller sample sizes the 95% confidence intervals of all provinces overlapped, limiting the ability to interpret levels of diversity. The rarefaction analysis based on 32 subtypes only identified a difference in subtype richness in the Atlantic region (Fig 3B), where a lower level of diversity was observed.

The presence of phylogeographic structure was tested using the program BaTS. When sequences were labeled according to province of origin, the parsimony score (PS) and association index (AI) statistics strongly rejected the null hypothesis of panmixis (Fig 4C). The monophyletic clade size statistic (MC) quantified the strength of each phylogeny-location association. With the exception of the Atlantic region, all regions demonstrated a significant association ($p < 0.05$). The MC statistics from British Columbia and Manitoba should be

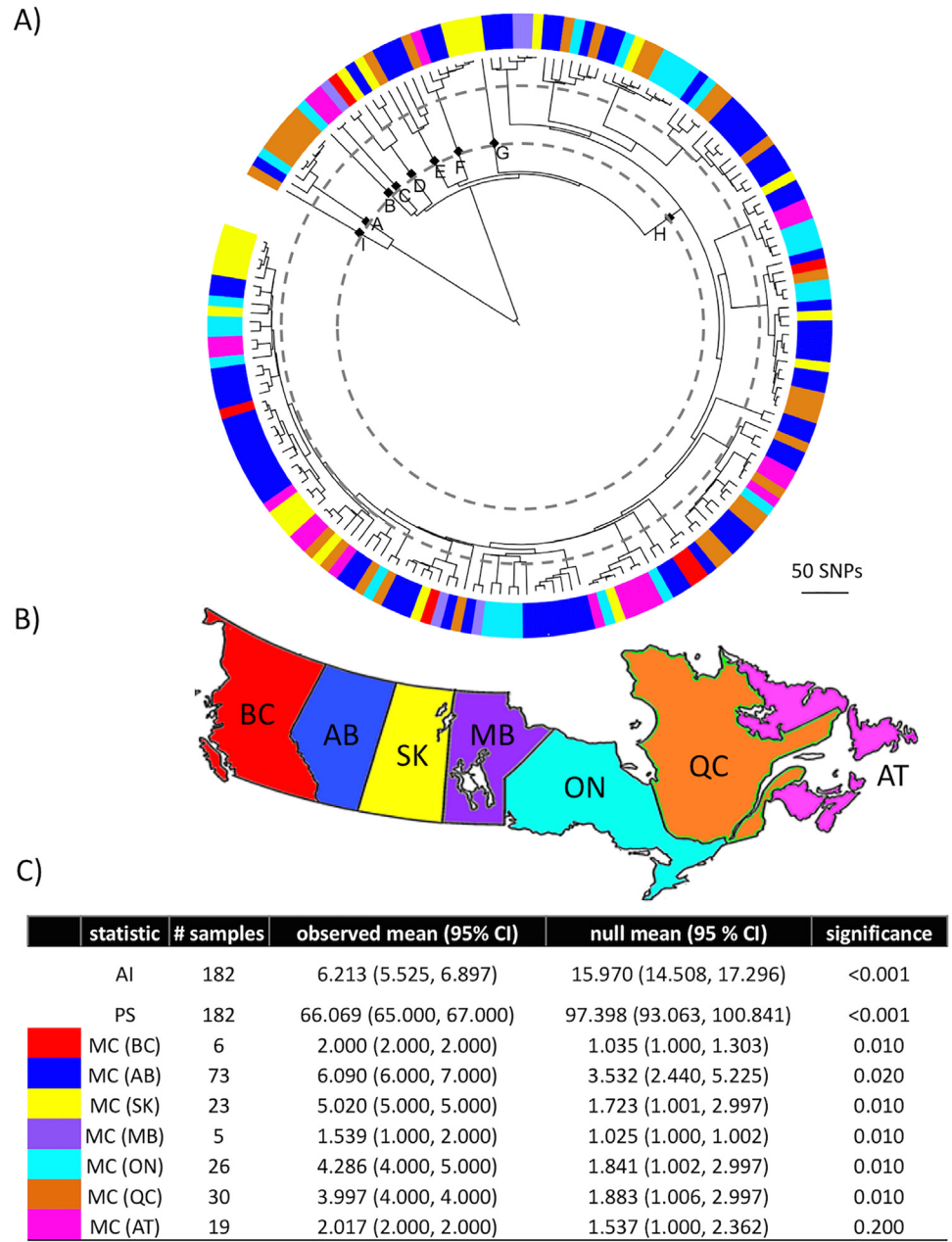


Fig 4. Phylogenetic clustering of MAP isolates among Canadian regions. A) Circularized maximum likelihood phylogenetic tree of 182 Canadian MAP isolates, rooted to the Type III isolate. Tips are colored according to province of origin (Alberta = blue, Ontario = light blue, British Columbia = red, Québec = orange, Saskatchewan = yellow, the Atlantic provinces = pink, Manitoba = purple). The branches leading to the nine subtypes are labeled with a square. Dotted lines indicate the threshold for the different subtype definitions used in the rarefaction analysis. B) Map of the 7 Canadian regions in which MAP isolates were obtained. C) The statistics (AI = association index, PS = parsimony score, MC = monophyletic clade size statistic), number of samples from each of the 7 regions, observed mean, null mean, and significance (p-value) are presented. Low AI and PS scores indicate a strong association, whereas high MC scores indicate a strong association.

doi:10.1371/journal.pone.0149017.g004

interpreted with caution due to their small sample size. A phylogenetic tree with the tips colored according to the province in which the isolate was sampled is depicted in Fig 4A. There

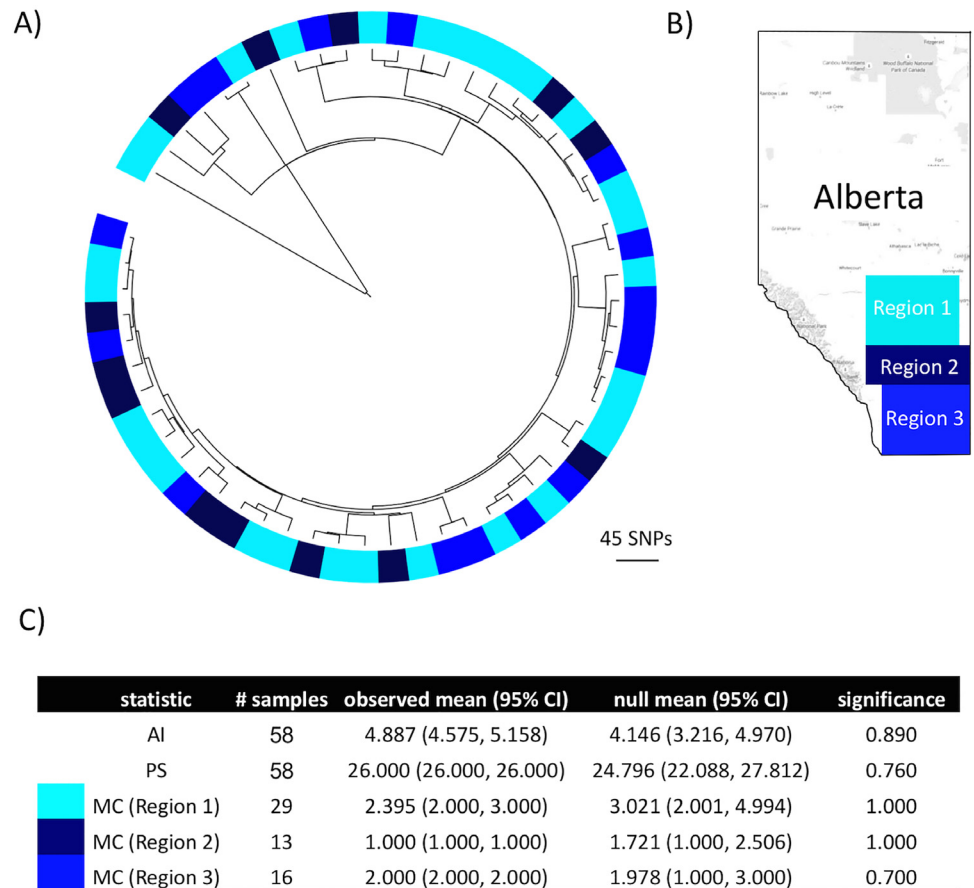


Fig 5. Phylogenetic clustering of MAP isolates in three regions in Alberta. A) Circularized maximum likelihood phylogenetic tree of 58 Alberta MAP isolates. Tips are colored according to the geographical location of originating farms (Region 1 = light blue, Region 2 = dark blue, Region 3 = blue). B) Map of Alberta with the three regions indicated by their respective colors. C) BaTS analysis of the phylogeny-region association. The statistics (AI = association index, PS = parsimony score, MC = monophyletic clade size statistic), number of samples from each region, observed mean, null mean, and significance (p-value) are presented. Low AI and PS scores indicate a strong association, whereas high MC scores indicate a strong association.

doi:10.1371/journal.pone.0149017.g005

were noticeably missing regions in subpopulations within the dominant clade. Isolates from the Atlantic region were not represented in the basal portion of the dominant clade (H) and isolates from Québec were not present in the distal portion of the dominant clade. An exact binomial test indicated these two regions were significantly underrepresented in those clades, with $p = 0.018$ and $p = 0.005$, respectively. When Alberta isolates were tested according to their location within the province, significant associations were not observed, as PS, AI, and all three MC statistics resulted in $p > 0.05$ (Fig 5).

Discussion

The WGS comparison of 182 Canadian MAP isolates originating from dairy herds in all 10 provinces to 26 diverse global isolates revealed the evolutionary relationship and subtype diversity of MAP. Although the proportion of subtypes could not be directly compared between the two datasets (because the global collection was selected to maximize genetic diversity), there

were subtypes in the global collection that were not found in Canada as well as global isolates that were interspersed with the Canadian isolates. This indicates that out of the known global diversity only a subset of subtypes has been introduced into Canada.

To our knowledge, this is the first report of Type III MAP isolated from a Canadian dairy farm. Furthermore, the identification of Type B isolates from three Canadian provinces, obtained through routine monitoring as part of JD control programs rather than through targeting of potential Type B sources, indicates that this strain type is more common than previously understood [12]. The close genetic relatedness of the USA Type B isolate [37] to the Canadian Type B isolates suggests a common ancestor within the past 25 years, given an estimated substitution rate of 0.3 SNPs per genome per year (Bryant et al., submitted for publication).

The dominant subtype included 86% of all Canadian MAP isolates. Multiple introduction events of this MAP subtype from other countries is likely, as five global isolates fell within the dominant clade and were often basal to the Canadian isolates. The large number of subpopulations within this type also supports this hypothesis. The current population structure of MAP in Canada was likely shaped by the importation of cattle from a limited number of countries (although uneven sampling makes it impossible to definitively identify the direction of transmission) and the spread of certain successful subtypes in the Canadian dairy environment. To further support the hypothesis that some subtypes may have a selective advantage, both *in vivo* and *in vitro* investigations into phenotypic differences, such as growth rates and invasion efficiencies, is warranted. Future vaccine development studies and molecular epidemiological analyses should also consider the relative frequencies of MAP subtypes in Canada with a focus on the dominant subtype.

The province-level diversity was assessed through rarefaction analyses. The greater diversity of divergent subtypes (using the strict definition) in Québec, which suggests a larger bacterial effective population size in this province, may be a reflection of their large number of dairy herds. In 2014, 5,894 dairy herds were operating in Québec, compared to just 566 in Alberta [38]. However, Saskatchewan, with only 166 dairy herds, exhibited greater MAP diversity than Ontario, with 3,926 herds, which doesn't fit the hypothesis that larger host population size and larger bacterial population size drive bacterial diversity. The rarefaction analysis with a more relaxed subtype definition indicated there is not a substantial difference in subtype diversity between provinces, with the exception of lower diversity seen in the Atlantic region. This complex distribution of MAP subtypes across the country has likely been shaped by herd management behaviors (such as biosecurity measures, JD testing strategy, cow-calf manure contact, colostrum management, etc.) that may select for certain MAP subtypes as well as the expansion of the Canadian dairy industry over the past two centuries, which may have contributed to repeated introductions.

The phylogeography was assessed at the between and within province level. The MC statistic was not significant in the Atlantic region, which includes four provinces separated by major geographical barriers. The grouping into one region may not be appropriate even though all four provinces were managed by the same JD control program [39]. There were large subpopulations within the dominant clade that did not include isolates from certain regions. This observation supports the rejection of the null hypothesis of panmixis at the between province level, even though there was still considerable overlap of subtypes across the country. There are several elements that could lead to an absence/low occurrence of subtypes in particular regions, such as local environmental factors, regional disease control strategies, a lack of introduction, or incomplete sampling. Animal trade is the probable driver of the shared subtypes between provinces. The movement of dairy cattle between herds in Canada is extensive but not well documented, as producers are not required to report livestock movements. A network analysis

of dairy cattle movements in Ontario between 2004 and 2006 estimated approximately 15% of dairy cattle shipments left the province (the majority going to Québec) and 10–20% remaining within the same postal code [40].

At the within-province level, MAP isolates from three regions in Alberta were not significantly correlated with the phylogeny. This is in contrast to *Mycobacterium bovis* in UK cattle, where geographic clustering of types at a smaller geographic scale is observed, and where wildlife populations are implicated as a potential reservoir for local and regional transmission [14,41]. In Alberta, however, the lack of spatial association suggests spatially-localized transmission mechanisms, such as wildlife, have not been a major driver of MAP transmission. Spatiotemporal clustering of one common MAP genotype was previously observed in Québec dairy herds [12], but genotypes were defined based on repetitive loci and may not have been a reliable measure of relatedness [8]. An understanding of dairy cattle movements within and between all provinces will help clarify the phylogeography of MAP in Canada, as the movement to and from major livestock markets may be influential in disseminating MAP. Modeling cattle contact networks may allow prediction of the patterns of herd-to-herd MAP transmission, similar to the approach used to estimate the main drivers of *M. bovis* transmission in British cattle [42] or human tuberculosis transmission in British Columbia, Canada [43].

The disparity in the number of MAP isolates sequenced from each province likely influenced the BaTS phylogeny-trait association. Manitoba and British Columbia had very small sample sizes, so the presence of just two neighboring tips in the phylogeny produced a significant signal. Additionally, while province is an important variable in the movement of livestock, there may be regional clusters that were not investigated in this study. For example, dairy farms in western Ontario may be more likely to buy/sell animals to dairy farms in Manitoba than eastern Ontario. This emphasizes the need for additional data regarding livestock movements in Canada. The number of shared subtypes with the USA is also an important research question; over the past 10 years, Canada has imported more than 200,000 dairy cattle from the United States [44]. This highlights the importance of assessing the population structure of MAP in the United States using WGS to appropriately evaluate the contribution of international MAP transmission.

Conclusions

Nine divergent MAP subtypes, including Type III and Type B, were identified in Canadian dairy herds, as well as a dominant subtype that comprised 86% of the Canadian isolates. The close similarity between the Type B isolates from the USA and Canada suggests a recent transmission event between bison and dairy cattle. The phylogenetic association of MAP isolates from different provinces indicates a degree of strain sharing within provinces, as well as inter-provincial movement of MAP subtypes. Further analysis of the population structure with a larger set of isolates would potentially identify which factors influence the degree of subtype diversity between different regions in Canada. SNPs alone are not sufficient to track MAP transmission given its slow growth rate and the extensive, undocumented interprovincial cattle movement. However, integrating genetic data with animal movement data will help clarify MAP transmission dynamics at a regional and national scale.

Supporting Information

S1 Fig. Likelihood mapping analysis of 182 Canadian MAP isolates. The phylogenetic signal of the dataset was tested by plotting 10,000 quartets, in which the unresolved quartets (10.8%) are shown in the central region of the triangle. 80.3% of the quartets were fully resolved and

8.8% were partially resolved.
(TIFF)

S1 Table. ID, origin, depth of coverage, and number of reads for each MAP isolate included in the study.
(XLSX)

Acknowledgments

The authors would like to thank Brian Radke of the British Columbia Ministry of Agriculture for providing MAP isolates from British Columbia, Robin King from Alberta Agriculture and Rural Development for providing historical MAP isolates from Alberta, and Uliana Kanevets for her help with MAP culture.

Author Contributions

Conceived and designed the experiments: HWB DH DFK GF OL GPK SLBM KT JDB. Performed the experiments: CA. Analyzed the data: CA HWB KS RNZ RK RB JDB. Contributed reagents/materials/analysis tools: DFK GF OL DH GPK SLBM KT HT. Wrote the paper: CA HWB JDB.

References

1. Barkema H. Global prevalence and economics of infection with *Mycobacterium avium* subsp. *paratuberculosis* in ruminants. *Paratuberculosis: Organism, Disease, Control*. 2010; Edited by Behr MA and Collins DM. Cambridge, MA: CAB International.
2. Behr MA. The path to Crohn's disease: is mucosal pathology a secondary event? *Inflamm Bowel Dis*. 2010; 16: 896–902. doi: [10.1002/ibd.21171](https://doi.org/10.1002/ibd.21171) PMID: [19924803](https://pubmed.ncbi.nlm.nih.gov/19924803/)
3. Barkema HW, Hendrick S, Buck JM de, Ghosh S, Kaplan GG, Rioux KP, et al. Zoonotic pathogens in the food chain. Wallingford: CABI; 2010. doi: [10.1079/9781845936815.0000](https://doi.org/10.1079/9781845936815.0000)
4. McKenna SLB, Keefe GP, Tiwari A, VanLeeuwen J, Barkema HW. Johne's disease in Canada part II: disease impacts, risk factors, and control programs for dairy producers. *Can Vet J*. 2006; 47: 1089–99. PMID: [17147140](https://pubmed.ncbi.nlm.nih.gov/17147140/)
5. Wolf R, Barkema HW, De Buck J, Slomp M, Flaig J, Hauptstein D, et al. High herd-level prevalence of *Mycobacterium avium* subspecies *paratuberculosis* in Western Canadian dairy farms, based on environmental sampling. *J Dairy Sci*. 2014; 97: 6250–9. doi: [10.3168/jds.2014-8101](https://doi.org/10.3168/jds.2014-8101) PMID: [25151881](https://pubmed.ncbi.nlm.nih.gov/25151881/)
6. Collins DM, Gabric DM, de Lisle GW. Identification of two groups of *Mycobacterium paratuberculosis* strains by restriction endonuclease analysis and DNA hybridization. *J Clin Microbiol*. 1990; 28: 1591–6. PMID: [2166089](https://pubmed.ncbi.nlm.nih.gov/2166089/)
7. Alexander DC, Turenne CY, Behr MA. Insertion and deletion events that define the pathogen *Mycobacterium avium* subsp. *paratuberculosis*. *J Bacteriol*. 2009; 191: 1018–25. doi: [10.1128/JB.01340-08](https://doi.org/10.1128/JB.01340-08) PMID: [19028885](https://pubmed.ncbi.nlm.nih.gov/19028885/)
8. Ahlstrom C, Barkema HW, Stevenson K, Zadoks RN, Biek R, Kao R, 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*. 2015; 16: 161. doi: [10.1186/s12864-015-1387-6](https://doi.org/10.1186/s12864-015-1387-6) PMID: [25765045](https://pubmed.ncbi.nlm.nih.gov/25765045/)
9. Amonsin A, Li LL, Zhang Q, Bannantine JP, Motiwala AS, Sreevatsan S, et al. Multilocus short sequence repeat sequencing approach for differentiating among *Mycobacterium avium* subsp. *paratuberculosis* strains. *J Clin Microbiol*. 2004; 42: 1694–702. doi: [10.1128/JCM.42.4.1694-1702.2004](https://doi.org/10.1128/JCM.42.4.1694-1702.2004) PMID: [15071027](https://pubmed.ncbi.nlm.nih.gov/15071027/)
10. Thibault VC, Grayon M, Boschiroli ML, Hubbans C, Overduin P, Stevenson K, 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. *J Clin Microbiol*. 2007; 45: 2404–10. doi: [10.1128/JCM.00476-07](https://doi.org/10.1128/JCM.00476-07) PMID: [17537942](https://pubmed.ncbi.nlm.nih.gov/17537942/)
11. Collins DM, Gabric DM de LG. Strain characterization of *Mycobacterium avium* subsp. *paratuberculosis*. *Paratuberculosis: Organism, Disease, Control*. 2010; Edited by Behr MA and Collins DM. Cambridge, MA: CAB International.

12. Sohal JS, Arsenault J, Labrecque O, Fairbrother J-H, Roy J-P, Fecteau G, 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*. 2014; 52: 2764–75. doi: [10.1128/JCM.00386-14](https://doi.org/10.1128/JCM.00386-14) PMID: [24829229](https://pubmed.ncbi.nlm.nih.gov/24829229/)
13. Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, Wagner DM. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect Genet Evol*. 2004; 4: 205–13. doi: [10.1016/j.meegid.2004.02.005](https://doi.org/10.1016/j.meegid.2004.02.005) PMID: [15450200](https://pubmed.ncbi.nlm.nih.gov/15450200/)
14. Biek R, O'Hare A, Wright D, Mallon T, McCormick C, Orton RJ, et al. Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathog*. 2012; 8: e1003008. doi: [10.1371/journal.ppat.1003008](https://doi.org/10.1371/journal.ppat.1003008) PMID: [23209404](https://pubmed.ncbi.nlm.nih.gov/23209404/)
15. Coscolla M, Gagneux S, Does M. tuberculosis genomic diversity explain disease diversity? *Drug Discov Today Dis Mech*. 2010; 7: e43–e59. doi: [10.1016/j.ddmec.2010.09.004](https://doi.org/10.1016/j.ddmec.2010.09.004) PMID: [21076640](https://pubmed.ncbi.nlm.nih.gov/21076640/)
16. Gollnick NS, Mitchell RM, Baumgart M, Janagama HK, Sreevatsan S, Schukken YH. Survival of *Mycobacterium avium* subsp. *paratuberculosis* in bovine monocyte-derived macrophages is not affected by host infection status but depends on the infecting bacterial genotype. *Vet Immunol Immunopathol*. 2007; 120: 93–105. doi: [10.1016/j.vetimm.2007.07.017](https://doi.org/10.1016/j.vetimm.2007.07.017) PMID: [17825427](https://pubmed.ncbi.nlm.nih.gov/17825427/)
17. Janagama HK, Jeong K il, Kapur V, Coussens P, Sreevatsan S. Cytokine responses of bovine macrophages to diverse clinical *Mycobacterium avium* subspecies *paratuberculosis* strains. *BMC Microbiol*. 2006; 6: 10. doi: [10.1186/1471-2180-6-10](https://doi.org/10.1186/1471-2180-6-10) PMID: [16478544](https://pubmed.ncbi.nlm.nih.gov/16478544/)
18. Motiwala AS, Janagama HK, Paustian ML, Zhu X, Bannantine JP, Kapur V, et al. Comparative transcriptional analysis of human macrophages exposed to animal and human isolates of *Mycobacterium avium* subspecies *paratuberculosis* with diverse genotypes. *Infect Immun*. 2006; 74: 6046–56. doi: [10.1128/IAI.00326-06](https://doi.org/10.1128/IAI.00326-06) PMID: [17057086](https://pubmed.ncbi.nlm.nih.gov/17057086/)
19. Borrmann E, Möbius P, Diller R, Köhler H. Divergent cytokine responses of macrophages to *Mycobacterium avium* subsp. *paratuberculosis* strains of Types II and III in a standardized *in vitro* model. *Vet Microbiol*. 2011; 152: 101–11. doi: [10.1016/j.vetmic.2011.04.002](https://doi.org/10.1016/j.vetmic.2011.04.002) PMID: [21555192](https://pubmed.ncbi.nlm.nih.gov/21555192/)
20. Lavers CJ, McKenna SLB, Dohoo IR, Barkema HW, Keefe GP. Evaluation of environmental fecal culture for *Mycobacterium avium* subspecies *paratuberculosis* detection in dairy herds and association with apparent within-herd prevalence. *Can Vet J*. 2013; 54: 1053–60. PMID: [24179240](https://pubmed.ncbi.nlm.nih.gov/24179240/)
21. Vary PH, Andersen PR, Green E, Hermon-Taylor J, McFadden JJ. Use of highly specific DNA probes and the polymerase chain reaction to detect *Mycobacterium paratuberculosis* in Johne's disease. *J Clin Microbiol*. 1990; 28: 933–7. PMID: [2351737](https://pubmed.ncbi.nlm.nih.gov/2351737/)
22. Ahlstrom C, Barkema HW, 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. *Appl Environ Microbiol*. 2014; 80: 534–9. doi: [10.1128/AEM.03212-13](https://doi.org/10.1128/AEM.03212-13) PMID: [24212568](https://pubmed.ncbi.nlm.nih.gov/24212568/)
23. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754–60. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
24. Li L, Bannantine JP, Zhang Q, Amonsin A, May BJ, Alt D, et al. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc Natl Acad Sci U S A*. 2005; 102: 12344–9. doi: [10.1073/pnas.0505662102](https://doi.org/10.1073/pnas.0505662102) PMID: [16116077](https://pubmed.ncbi.nlm.nih.gov/16116077/)
25. Wynne JW, Seemann T, Bulach DM, Coutts SA, Talaat AM, Michalski WP. Resequencing the *Mycobacterium avium* subsp. *paratuberculosis* K10 genome: improved annotation and revised genome sequence. *J Bacteriol*. 2010; 192: 6319–20. doi: [10.1128/JB.00972-10](https://doi.org/10.1128/JB.00972-10) PMID: [20870759](https://pubmed.ncbi.nlm.nih.gov/20870759/)
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–9. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
27. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003; 52: 696–704. PMID: [14530136](https://pubmed.ncbi.nlm.nih.gov/14530136/)
28. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 2012; 9: 772. doi: [10.1038/nmeth.2109](https://doi.org/10.1038/nmeth.2109)
29. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*. 2011; 39: W475–8. doi: [10.1093/nar/gkr201](https://doi.org/10.1093/nar/gkr201) PMID: [21470960](https://pubmed.ncbi.nlm.nih.gov/21470960/)
30. Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 2002; 18: 502–4. doi: [10.1093/bioinformatics/18.3.502](https://doi.org/10.1093/bioinformatics/18.3.502) PMID: [11934758](https://pubmed.ncbi.nlm.nih.gov/11934758/)
31. Strimmer K, von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc Natl Acad Sci USA*. 1997; 94: 6815–9. PMID: [9192648](https://pubmed.ncbi.nlm.nih.gov/9192648/)
32. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. *Genetics*. 2006; 172: 2665–81. doi: [10.1534/genetics.105.048975](https://doi.org/10.1534/genetics.105.048975) PMID: [16489234](https://pubmed.ncbi.nlm.nih.gov/16489234/)

33. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006; 23: 254–67. doi: [10.1093/molbev/msj030](https://doi.org/10.1093/molbev/msj030) PMID: [16221896](https://pubmed.ncbi.nlm.nih.gov/16221896/)
34. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org/>.
35. Parker J, Rambaut A, Pybus OG. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol.* 2008; 8: 239–46. doi: [10.1016/j.meegid.2007.08.001](https://doi.org/10.1016/j.meegid.2007.08.001) PMID: [17921073](https://pubmed.ncbi.nlm.nih.gov/17921073/)
36. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 2012; 29: 1969–73. doi: [10.1093/molbev/mss075](https://doi.org/10.1093/molbev/mss075) PMID: [22367748](https://pubmed.ncbi.nlm.nih.gov/22367748/)
37. Whittington RJ, Marsh IB, Whitlock RH. 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. *Mol Cell Probes.* 2001; 15: 139–45. doi: [10.1006/mcpr.2001.0346](https://doi.org/10.1006/mcpr.2001.0346) PMID: [11352594](https://pubmed.ncbi.nlm.nih.gov/11352594/)
38. Animal Industry Market Information System dairy genetics database. 2015. Available: http://www.dairyinfo.gc.ca/index_e.php?s1=df-fcil&s2=farm-ferme&s3=nb. Accessed 6 June 2015.
39. Atlantic Johnes Disease Initiative. 2015. Available: <http://www.atlanticjohnes.ca>. Accessed 6 June 2015.
40. Dubé C, Ribble C, Kelton D, McNab B. Comparing network analysis measures to determine potential epidemic size of highly contagious exotic diseases in fragmented monthly networks of dairy cattle movements in Ontario, Canada. *Transbound Emerg Dis.* 2008; 55: 382–92. doi: [10.1111/j.1865-1682.2008.01053.x](https://doi.org/10.1111/j.1865-1682.2008.01053.x) PMID: [18840200](https://pubmed.ncbi.nlm.nih.gov/18840200/)
41. Donnelly CA, Nouvellet P. The contribution of badgers to confirmed tuberculosis in cattle in high-incidence areas in England. *PLoS Curr.* 2013; 5. doi: [10.1371/currents.outbreaks.097a904d3f3619db2fe78d24bc776098](https://doi.org/10.1371/currents.outbreaks.097a904d3f3619db2fe78d24bc776098)
42. Green DM, Kiss IZ, Mitchell AP, Kao RR. Estimates for local and movement-based transmission of bovine tuberculosis in British cattle. *Proc Biol Sci. The Royal Society*; 2008; 275: 1001–5. doi: [10.1098/rspb.2007.1601](https://doi.org/10.1098/rspb.2007.1601) PMID: [18252669](https://pubmed.ncbi.nlm.nih.gov/18252669/)
43. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med.* 2011; 364:730–9. doi: [10.1056/NEJMoa1003176](https://doi.org/10.1056/NEJMoa1003176) PMID: [21345102](https://pubmed.ncbi.nlm.nih.gov/21345102/)
44. Animal Industry Market Information System trade database. 2013. Available: http://www.dairyinfo.gc.ca/index_e.php?s1=df-fcil&s2=imp-exp&s3=gen. Accessed 6 June 2015.