



Article First Report and Comparative Genomic Analysis of Mycoplasma capricolum subsp. capricolum HN-B in Hainan Island, China

Zhenxing Zhang ^(D), Meirong He, Junming Jiang, Xubo Li, Haoyang Li, Wencan Zhang, Si Chen ^(D), Li Du, Churiga Man, Qiaoling Chen ^{*(D)}, Hongyan Gao ^{*} and Fengyang Wang ^{*}

Hainan Key Laboratory of Tropical Animal Reproduction & Breeding and Epidemic Disease Research, Engineering Key Laboratory of Haikou, College of Animal Science and Technology, Hainan University, Haikou 570228, China

* Correspondence: chenqiaoling@hainanu.edu.cn (Q.C.); gaohongyan@hainanu.edu.cn (H.G.); 990805@hainanu.edu.cn (F.W.)

Abstract: *Mycoplasma capricolum* subsp. *Capricolum* (Mcc) is an important member of the *Mycoplasma mycoides* cluster (Mm cluster) and causes caprine contagious agalactia. Mcc can infect goats of all age groups, especially pregnant ewes and kids. It can cause the abortion in pregnant ewes and the death of goat kids, leading to enormous losses in the goat breeding industry. To date, the prevalence of epidemic Mcc strains on Hainan Island, China, remains unclear. This study aimed to isolate and identify Mcc strains endemic to Hainan Island, China. Genome sequencing and comparative genomic analysis were performed to reveal the molecular characteristics and evolutionary relationships of the isolated strain. Mcc HN-B was isolated and identified in Hainan Island, China. The Mcc HN-B genome consists of a 1,117,925 bp circular chromosome with a 23.79% G + C content. It contains 912 encoding genes, 3 gene islands, and 14 potential virulence genes. The core genome with the features of the Mm cluster and the specific genes of Mcc HN-B were identified by comparative genomic analysis. These results revealed the evolutionary relationship between Mcc HN-B and other members of the Mm cluster. Our findings provide a reference for further studies on the pathogenic mechanism and local vaccine development of Mcc.

Keywords: *Mycoplasma capricolum* subsp. *capricolum; Mycoplasma mycoides* cluster; gene islands; virulence factor; Hainan Island; comparative genomics

1. Introduction

The *Mycoplasma mycoides* cluster (Mm cluster) is a group of several mycoplasmas that have close genetic relationships and the same pathogenic characteristics [1,2]. It includes pathogenic mycoplasmas, such as *Mycoplasma mycoides* subsp. *capri* (Mmc), *Mycoplasma mycoides* subsp. *mycoides* large colony (MmmLC), *Mycoplasma capricolum* subsp. *capricolum* (Mcc), *Mycoplasma capricolum* subsp. *capripneumoniae* (Mccp), *Mycoplasma mycoides* subsp. *mycoides* small colony (MmmSC), and *Mycoplasma leachii* (MI) [3]. Mccp is the causative agent of contagious caprine pleuropneumonia, which can lead to huge economic losses in the goat breeding industry [4,5]. MmmSC can cause contagious bovine pleuropneumonia, presenting an enormous threat to the cattle industry [6]. Mcc, Mmc, and MmmLC can cause mastitis, arthritis, keratoconjunctivitis, pneumonia, and septicemia (MAKePS) syndrome in ruminants and are also the causative agents of contagious agalactia [4,7]. In addition, Mcc is highly virulent and pathogenic in goats. It can infect goats of all age groups and cause abortion of ewes and death of cubs [8]. Nicolas et al. [9] found that MmmLC and Mcc can be interspecifically transmitted between different taxa of ruminants, which may cause great losses in the breeding industry.

With the development of bacterial isolation and culture and molecular diagnostic techniques, Mcc has been found in many countries and regions, such as Spain, Portugal,



Citation: Zhang, Z.; He, M.; Jiang, J.; Li, X.; Li, H.; Zhang, W.; Chen, S.; Du, L.; Man, C.; Chen, Q.; et al. First Report and Comparative Genomic Analysis of *Mycoplasma capricolum* subsp. *capricolum* HN-B in Hainan Island, China. *Microorganisms* **2022**, *10*, 2298. https://doi.org/10.3390/ microorganisms10112298

Academic Editors: Rohini Chopra-Dewasthaly and Lucas Miranda Marques

Received: 2 October 2022 Accepted: 16 November 2022 Published: 19 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). and the United States [8,10,11]. Researchers have used various methods to study Mcc to prevent and treat Mcc-related diseases. Using PCR and restriction enzyme digestion technologies, Rodriguez et al. [12] established an approach to differentiate Mcc from other members of the Mm cluster. Based on the suppression subtractive hybridization technology, Maigre et al. [13] developed a specific PCR diagnostic method for Mcc. In addition, comparative genomic analysis has been performed to reveal the genetic relationship and variation between different strains of Mcc and between Mcc and other mycoplasmas [14,15]. Moreover, antimicrobial susceptibility tests of Mcc have been carried out, and further studies have investigated its drug resistance mechanisms [16,17].

To date, a total of five Mcc genomes can be found in the National Center for Biotechnology Information (NCBI, 24 September 2022), yet only one genome has been completely assembled. The complete genome of Mcc-endemic strains is a fundamental prerequisite for studying their molecular characteristics and pathogenic mechanisms. It has an important reference value for Mcc evolutionary relationship research, epidemiological investigations, and regional vaccine development.

In the present study, Mcc HN-B was isolated and identified for the first time on Hainan Island, China. We obtained a high-quality Mcc HN-B genome using whole-genome sequencing. All members of the Mm cluster (Mmc HN-A, MmmLC 95010, Mcc ATCC 27343, Mccp M1601, MmmSC PG1, and Ml PG50) were selected to perform comparative genomic analysis with Mcc HN-B. This study provides a valuable reference for acquiring strain information on Mcc, analyzing the molecular characteristics of the Mcc HN-B genome in detail, and revealing the genetic evolutionary relationships among all members of the Mm cluster. In addition, it will help foster regional prevention, pathogenic mechanism research, and drug development for Mcc.

2. Materials and Methods

2.1. Isolation and Identification of Mcc HN-B

In a black goat breeding factory, a two-month-old female lamb suffered from elbow swelling of the left forelimb, limp, ataxia, and other symptoms. A sterile syringe was used to draw the liquid inside the cyst of the elbow joint to isolate and identify the bacteria. The volume of the joint fluid was approximately 100 μ L, which was clear with a light-yellow clot. It was stored by the Hainan Key Laboratory of Tropical Animal Breeding and Disease Research.

We added 50 μ L of joint fluid to 450 μ L of sterile PBS and mixed thoroughly. This procedure was repeated thrice. Then, different concentrations of 100 μ L of articular fluid were inoculated into 10 mL of mycoplasma liquid medium, which included 0.03 g glucose, 0.025 g yeast powder, 0.02 g sodium pyruvate, 0.21 g pleuropneumonia-like organism (PPLO) broth (Difco, Tucker, GA, USA), 8000 IU penicillin G (BioFroxx, Einhausen, Germany), 2 mL horse serum (Solarbio, Beijing, China), and 25 μ L of 0.4% phenol red solution (Amresco, Boise, ID, USA). In addition, 1.5% agar powder (BioFroxx, Einhausen, Germany) was added to the mycoplasma liquid medium to prepare a mycoplasma solid medium. After inoculation, the medium was placed in a cell incubator (37 °C and 5% CO₂) for culturing.

Changes in the mycoplasma liquid and solid media were observed and recorded. When the liquid medium became yellow or cloudy, the medium was filtered through a filter with a pore size of 0.45 μ m. The filtered medium was seeded and subcultured in fresh medium at 10% volume. Purified bacteria were obtained by repeated passaging (three times). The inoculated mycoplasma solid medium was observed daily under an ordinary light microscope. Once suspicious colonies were found, they were cloned and purified on new solid medium over time. Giemsa staining was performed, and the cells were observed and photographed under a microscope.

The synthetic universal primers for the bacterial *16S rRNA* gene [18] were F, 5'-AGAGTTTGATCCTGGCTCAG-3' and R, 5'-GGTTACCTTGTTACGACTT-3'. Purified colonies were selected for amplification and identification of the *16S rRNA* gene. The PCR

system included upstream and downstream primers (20 pmol each), ddH₂O (20 μ L), 2× Taq PCR Mix (25 μ L), and 1 μ L of DNA template. The PCR procedure involved predenaturation at 94 °C for 5 min, denaturation at 94 °C for 30 s, annealing at 56 °C for 50 s, extension at 72 °C for 90 s, 32 cycles of circulation, and termination of extension at 72 °C for 10 min.

2.2. Mcc HN-B Strain Culture and Genomic DNA Extraction

Purified colonies were selected and inoculated into 100 mL of mycoplasma liquid medium for expanded culture. The genome of Mcc HN-B was extracted using a HiPure Bacterial DNA Kit (Magen, Guangzhou, China). Qubit (Thermo Fisher Scientific, Waltham, MA, USA) and NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) were used to test the DNA quality.

2.3. ONT Sequencing and Illumina Sequencing

After the DNA quality test, DNA libraries for Oxford Nanopore Technology (ONT) sequencing and Illumina sequencing were constructed. First, according to the protocol offered by ONT, the genomic DNA was broken into segments averaging approximately 10 kB using G-tubes (Covaris, Woburn, MA, USA). Next, NEBNext End Repair/dA-Tailing Module reagents (E7546, NEB, USA) and the NEBNext FFPE DNA Repair kit (M6630, NEB, USA) were used for end repair. NEBNext Quick Ligation Module reagents (E6056, NEB, USA) were used for joint connection. Finally, we used the AMPure XP system (Beckman Coulter, Brea, CA, USA) to purify the DNA library and performed sequencing using the long-read sequencing platform Oxford Nanopore PromethION (Oxford Nanopore Technologies, Oxford, UK).

In accordance with the instructions of the NEBNext[®] MLtraTM DNA Library Prep Kit for Illumina (NEB, USA), the genomic DNA segmented by random ultrasound was subjected to end repair, polyA tail processing, and adaptor ligation. Then, the PCR and AMPure XP systems (Beckman Coulter, Brea, CA, USA) were employed for enrichment and purification of the library. After qualification testing using a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA), paired-end technology (PE 150) was applied for sequencing on an Illumina NovaSeq 6000 sequencer.

2.4. Genome Assembly and Genome Component Annotation

Flye (version 2.8.1-b1676) [19] was used to acquire long reads from the ONT sequencing and assemble from scratch. FASTP (version 0.20.0) [20] was used to filter the original data from the Illumina platform. Finally, Pilon (version 1.23) [21] was used to correct the genome sequence via filtered reads and obtain the final genome sequence.

The Mcc HN-B composition prediction involves open reading frames (ORFs), rRNAs, tRNAs, sRNAs, gene islands (GIs), clustered regularly interspaced short palindromic repeats (CRISPR), transposons, interspersed repeat elements, tandems, and prophages. They were predicted using the NCBI prokaryotic genome annotation pipeline [22], RNAmmer (version 1.2) [23], tRNAscan (version 1.3.1) [24], cmscan (version 1.1.2) [25], IslandPath-DIMOB (version 1.0.0) [26], CRISPRFinder (4.2.17) [27], TransposonPSI (version: 20100822) [28], RepeatMasker (version 4.0.5) [29], TRF (version 4.09) [30], and PHAST (version 2.0) [31], respectively.

2.5. Gene Function Analysis

Based on the screening standard of E-value $< 1 \times 10^{-5}$, databases, such as nonredundant protein sequence database (Nr), UniProt/Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COG), Carbohydrate-Active EnZYmes database (CAZy), Transporter Classification Database (TCDB), Pathogen–Host Interactions database (PHI), and Virulence Factors of Pathogenic Bacteria database (VFDB), were used to analyze the genomic sequence of Mcc HN-B in a comparative manner and annotate the gene function. According to the default software parameters, the Gene Ontology database (GO), Pfam database (version 32.0) and Comprehensive Antibiotic Resistance Database (CARD) were used to predict DNA functions. The software information is listed in Table S1.

In addition, secreted and transmembrane proteins were predicted using SignalP4.0 [32] and TMHMM (version 2.0) [33], respectively. Type three secretion systems (T3SSs) and secondary metabolic gene clusters were predicted using EffectiveT3 (version 1.0.1) [34] and antiSMASH (version 4.1.0) [35]. We also predicted the two-component systems (TCSs) of Mcc HN-B.

2.6. Collinearity Analysis and SNP/InDel/SV Statistics

As the target genome, Mcc HN-B belongs to one of the members of the Mm cluster. The six genomes selected were different subspecies within the Mm cluster. First, MUMmer (version 3.1) [36] was used to compare the target and reference genomes to identify the large-scale collinearity between genomes. Subsequently, SyRI (version 1.4) [37] was applied to test the local position arrangement of the samples.

Single-nucleotide polymorphisms (SNPs) and insertions–deletions (InDels) were tested and analyzed using MUMmer (version 3.1). Structural variation (SV) was examined and analyzed using SyRI (version 1.4).

2.7. Gene Family Analysis

In accordance with the bidirectional best-hit standard (80% of the shortest protein sequence has 40% amino acid similarity), gene family analysis was performed on the target and reference genomes. Diamond [38] (version 2.0.7) and OrthoMCL (version 1.4) [39] were used to perform a comparison and similarity clustering of their amino acid (or nucleotide) sequences. Finally, homologous gene clusters and species distributions of each protein cluster were obtained.

2.8. Core/Pan-Genome Analysis

The pan-genome includes both core and dispensable genomes. Among them, the genes present in all strains constitute the core genome, and genes excluding these common genes constitute the dispensable genome. In dispensable genomes, genes that are unique to only one strain are named 'unique genes'; a gene that is shared by more than one strain, but not all, is named an 'accessory gene'.

Core/pan-genome analysis was performed using Mcc HN-B and the six reference genomes. Strain-specific genes of Mcc HN-B and reference strains were counted according to the gene family clustering results in Section 2.7. Strain-specific genes included genes in strain-specific gene families and genes that were not involved in the above clustering.

2.9. Phylogenetic Tree Analysis and ANI Analysis

Considering the homologous gene cluster analysis results of Mcc HN-B and the reference genomes, we selected homologous genes with single copies. Multiple sequence alignment and alignment quality control were performed using MUSCLE [40] (version 3.8.31) and Gblocks [41] (version 0.91B), respectively. IQ-Tree [42] (version 1.6.3) was used to construct phylogenetic trees according to the maximum likelihood method.

The average nucleotide identity (ANI) value refers to the average base similarity between homologous segments of two microbial genomes. Normally, an ANI value of 95% is used as the classification threshold to distinguish between different species, which shows a high degree of discrimination among closely related species. We used Pyani [43] (version 0.2.7) to calculate the ANI value of the alignment region between the Mcc HN-B genome and each reference genome.

2.10. GenBank Accession Number

The annotated whole-genome sequence of the Mcc HN-B strain was submitted to the GenBank database under accession number CP101903.

3. Results and Discussion

3.1. Results of Mcc HN-B Isolation and Identification

The joint fluid was cultured in mycoplasma liquid medium. After 17 h, the color of the liquid medium changed from red to yellow. After 25 h, the liquid medium became turbid and precipitated. After 48 h, the color of the liquid medium did not change, and the precipitation reached its maximum. Simultaneously, the joint fluid was inoculated into the mycoplasma solid medium. Small transparent colonies began to appear 24 h after inoculation. After 36 h, the colonies continued to grow, and the color became milky white. After three subcultures, we successfully isolated and purified the mycoplasma strain.

Under the low-power microscope, the colonies were mostly round, and the central umbilicuses were obvious (Figure 1A). Single colonies of this strain were selected for Giemsa staining and observed under magnification of 1000 times (Figure 1B); purple particles were observed in the microscopic field, and the isolated strains were spherical or arc shaped.



Figure 1. Mcc HN-B isolation and identification. (**A**) The morphology of the isolated strain, magnified 40 times (scale bar = 250 μ m); (**B**) Giemsa staining results of the isolated strain, magnified 1000 times (scale bar = 10 μ m): the top left picture results from the red-framed area magnified three times. (**C**) Agarose gel electrophoresis results of the *16S rRNA* gene: (M) D2000 Marker, (Line 1) the isolated strain, (Line 2) blank control.

The *16S rRNA* gene of the isolated strain was successfully amplified and sequenced (Figure 1C). The NCBI analysis results showed that the *16S rRNA* gene sequence of the isolated strain had the highest homology (99.71%) with Mcc 14DD0024. The strain was identified as Mcc by culture characteristics, morphological characteristics, and PCR amplification, and named "Mcc HN-B."

3.2. General Characteristics of the Mcc HN-B Genome

To ensure the accuracy and reliability of the subsequent information analysis, the original data generated by the ONT and Illumina sequencing platforms were filtered and

processed to obtain valid data. The Illumina sequencing data and ONT sequencing data of Mcc HN-B were 1.6 Gb (\sim 1428.57 \times) and 1.12 Gb (\sim 1000 \times), respectively. Through genome assembly and correction, the complete genomic sequence of Mcc HN-B, with a total length of 1,117,925 bp, was obtained. Combined with the prediction results of the coding genes, the Circos software (version 0.69–9) was used to draw the genome circle map of Mcc HN-B to comprehensively display its genome characteristics (Figure 2).



Figure 2. Whole-genome circle map of Mcc HN-B. The circles' diagram from outside to inside is plus and minus strand gene (gray), plus and minus strand COG, plus and minus strand KEGG, plus and minus strand GO, ncRNA, GC content, and GC skew. KEGG uses the topmost category, represented by five colors. Genes that correspond to more than one category are assigned the sixth color. GC and GC skew are positive outward and negative inward, respectively. The window and step size are both 1000.

The whole genome of the Mcc HN-B strain with a 23.79% G + C content includes 912 coding genes, 3 GIs, 175 tandem repeats, and 1 DNA transposon (Table 1). We first predicted six potential CRISPRs in Mcc HN-B (Table S2), which require further validation. Additionally, 912 proteins were encoded by the Mcc HN-B genome, including 12 transmembrane proteins (Table S3) and 46 secreted proteins (Table S4). The results of the functional annotation of the Mcc HN-B genome are shown in Table S5.

Item	Number	Item	Number
Genome size (bp)	1,117,925	Number of rRNA genes	6
Genome GC content (%)	23.79	Number of tRNA genes	31
Number of genes	912	Number of sRNA genes	1
Length of gene (bp)	45-5895	Number of interspersed repeats	6
Total length of gene (bp)	983,931	Number of short interspersed elements	4
Average length of gene (bp)	1078.87	Number of long interspersed repeated sequences	1
GC content of gene region (%)	23.95	Number of DNA elements	1
Total length of gene/genome (%)	88.01	Number of tandem repeats	175
Total length of intergenic region/genome (%)	11.99	Total bases in tandem repeats	20,145
GIs number	3	Number of prophages	0
Total GI length (bp)	69,172	Average length (bp)	23,057.33

Table 1. General characteristics of the Mcc HN-B genome.

3.2.1. GIs

Three GIs were first identified in the Mcc HN-B genome. Among them, 9, 32, and 26 genes were included in GI1 (GI_1), GI2 (GI_2), and GI3 (GI_3), respectively (Figure 3, Table S6).



Figure 3. Gene island composition of Mcc HN-B. Note: The abscissa represents the length scale.

On GI_1, NO343_00050, NO343_00060, and NO343_00070 encode three membrane proteins; NO343_00075 encodes one lipoprotein; and NO343_00055, NO343_00080, and NO343_00085 encode three hypothetical proteins. The DUF285 family protein encoded by NO343_00055 may be related to the adhesion of Mcc HN-B. NO343_00065 encodes a DnaJ-domain-containing protein [44]. It could act as an auxiliary molecular chaperone of heat shock protein 70 (HSP70) to participate in different physiological processes. NO343_00090 encodes a DDE transposase, which may provide a basis for inserting sequence transposition activity and promoting the evolution and environmental adaptation of bacteria [45].

GI_2 and GI_3 had similar compositions. Among them, 20 genes encode proteins that are structurally and functionally similar and arranged in highly similar positions. These genes encode hypothetical proteins, membrane proteins, and lipoproteins that are related to the virulence of Mcc HN-B. *NO343_02020* in GI_2 and *NO343_04000* in GI_3 encode transposases of the IS3 family. We speculated that GI_2 and GI_3 have undergone gene transfer in related species or the same species as Mcc, eventually leading to their similar structure and genetic composition.

Additionally, compared with GI_3, GI_2 encodes more proteins, including three ATP-binding proteins (NO343_02160, NO343_02165, and NO343_02170), two membrane proteins (NO343_02145 and NO343_02150), three hypothetical proteins (NO343_02135, NO343_02155, and NO343_02175), and one TIR-domain-containing protein (NO343_02015).

Comparatively, GI_3 encodes an extra hypothetical protein (NO343_04030) and two membrane proteins (NO343_04020 and NO343_04025).

3.2.2. Virulence Factors

VFDB is a database dedicated to the study of pathogenic factors in bacteria, chlamydia, and mycoplasma. Fourteen genes in the Mcc HN-B genome were annotated to the VFDB database. Among them, six and three genes were annotated to surface lipoproteins and capsules, respectively (Table S7). *NO343_00170*, *NO343_03810*, *NO343_04655*, *NO343_04660*, *NO343_04665*, and *NO343_04670* encode six Vmm proteins, which are all surface lipoproteins, contributing to high-frequency phase variation in immunodominant antigens. The proteins encoded by *GalU* (*NO343_02230*), *OppF* (*NO343_02495*), and *MSC_0991* (*NO343_02500*) belong to the capsule. Only one gene was annotated to GAPDH, hemolysin, elongation factor thermal unstable (EF-Tu), streptococcal enolase, or pyruvate dehydrogenase E1 beta subunit (PDH-B).

gapA (*NO343_00195*) encodes the GAPDH protein of Mcc HN-B. Recent studies [46] have indicated that the GAPDH protein is not only a key enzyme in the glycolytic metabolic pathway but also involved in intracellular digestion, DNA repair, apoptosis, and other life processes. Hoelzle et al. [47] reported that the MSG1 protein with GAPDH activity in *Mycoplasma suis* (*M. suis*) also functions in adhesion to porcine erythrocytes. We speculated that NO343_00195 plays an adhesive role in the Mcc HN-B strain to enhance Mcc infection in the host.

Hemolysin is a common exotoxin that can cause the dissolution of red blood cells, leading to capillary necrosis [48]. *hlyA* often exists in *Listeria monocytogenes* [49] and *Vibrio cholerae* [50]. Additionally, some pathogenic mycoplasma genomes contain genes encoding hemolysin A, such as *Mycoplasma hyopneumoniae* (*M. hyopneumoniae*), *Mycoplasma ovipneumoniae*, Mccp [51], and *Mycoplasma conjunctivae*. In the present research, we found that *hlyA* (*NO343_02185*) also exists in the Mcc HN-B genome, with a sequence homology rate of 98.02% compared with *hlyA* (*MCAP_0055*) in the Mcc ATCC 27343 genome submitted in 2005.

PDHB is an important component of pyruvate dehydrogenase. EF-Tu is an indispensable functional protein for life activities, accounting for approximately 10% of the total protein of *Mycoplasma pneumoniae* (*M. pneumoniae*) [52]. Pinto et al. [53] identified EF-Tu and PDHB as highly antigenic proteins of *M. hyopneumoniae* by Western blotting, which can be recognized by the host immune system. Dallo et al. [54] demonstrated that EF-Tu and PDH-B, two surface proteins of *M. pneumoniae*, could bind to fibronectin, which is a component of extracellular matrix, and affect mycoplasma adherence to the host and infection [55]. *pdhB* (*NO343_02990*) and *tuf* (*NO343_02670*) were identified in the Mcc HN-B genome, which might also facilitate the adhesion and infection of Mcc to the host.

 α -enolase exists on the surface of many bacteria, has good immunogenicity in *M. suis*, and can enhance the adhesion ability of bacteria to the host [56]. Esgleas et al. showed that α -enolase is a surface protein with fibronectin-binding activity that plays a role in bacterial adhesion and invasion [57,58]. The potential virulence gene *eno* (*NO343_02925*) was also identified in Mcc HN-B, which may be important for Mcc to invade the host.

We also found that *NO343_03060* and *NO343_03295* encode two kinds of peptidase S41 in the Mcc HN-B genome. A study demonstrated [59] that S41 peptidase is closely related to the proteolytic phenotype of Mmc. Deletion of the S41 peptidase can cause the phenotypic change in Mmc and increase the production of hydrogen peroxide. Therefore, NO343_03060 and NO343_03295 may be vital for the proteolytic phenotype and virulence sensitivity of Mcc HN-B.

Finally, a type I restriction modification (RM) system was identified in the Mcc HN-B genome. HsdR is encoded by *NO343_01990*, HsdM by *NO343_01975*, and HsdS by *hsdS* (*NO343_01985*, *NO343_02000*, *NO343_04805*, *NO343_04810*). This system contains repeated *hsdS* genes, and Mcc HN-B may change its phenotype through phase variation [60]. Meanwhile, the type I RM system may influence the virulence of Mcc HN-B by regulating

the expression of virulence genes [61,62]. *NO343_01175* and *NO343_03390* of Mcc HN-B encoding two type II RM system-related proteins were also identified.

3.2.3. Metabolism

The histidine phosphocarrier protein (HPr) is a component of the phosphotransferase system (PTS). Kundig [63] et al. reported its function in hexose phosphorylation in 1964. The PTS system has been found in an increasing number of fungi, archaea, and bacteria. Most of the known PTS systems have similar structures, consisting of the enzyme I complex (EI), HPr, and enzyme II complex (EIIA, EIIB, EIIC, EIID). The main function of the PTS system is to mediate the absorption and phosphorylation of carbohydrates and to participate in the regulation of carbon metabolism and bacterial virulence [64].

In this study, we identified 20 Mcc HN-B genes involved in the PTS system. Two cytosolic phosphotransferases, EI and HPr, are encoded by NO343_03025 and NO343_00615, respectively. NO343_03030 encodes EIIA^{Glc}. NO343_01255, NO343_01260, NO343_01275, NO343_01560, and NO343_02375 encode EIICB^{Glc}. NO343_01870 and NO343_03925 encode EIIA^{Mtl}. NO343_01860 and NO343_03920 encode EIICB^{Mtl}. NO343_01610 encodes EIIC^{Lev}. NO343_04645 encodes an ascorbate PTS system EIIB component. NO343_02540 encodes the ascorbate PTS system EIICBA or EIICB component. NO343_02675 encodes the sucrose PTS system EIIBCA or EIIBC component. NO343_04800 encode 1-phosphofructokinase.

Ten different types of EII were identified in Mcc HN-B. These enzymes enable the PTS system of Mcc HN-B to transport a wide range of carbohydrates. The major glucose transport system in Mcc HN-B was the glucose PTS system. Mcc also contained mannitol PTS, *N*-acetylglucosamine PTS, sucrose PTS, and fructose PTS. Studies have shown that *N*-acetylglucosamine PTS is more effective in promoting glucose transport and utilization than maltose PTS and β -glucoside PTS in the absence of glucose PTS, mannose PTS, and glucokinase [65]. Christine et al. proved that when fructose is the only carbon source, the fructose PTS of *Clostridium acetobutylicum* becomes the main fructose uptake system [66]. Combined with the PTS characteristics of Mcc HN-B, we hypothesized that the *N*-acetyl glucosamine PTS could act as a substitute to promote glucose transport when Mcc HN-B lacks glucose PTS and glucokinase. Moreover, the presence of fructose PTS in Mcc HN-B may make it possible to use fructose as a carbon source, thus enhancing the environmental stress resistance of the strain.

Additionally, 33 Mcc HN-B genes were involved in the metabolic process of the ATPbinding cassette (ABC) transporters, including a general nucleoside transport system, a phosphonate transport system, and an oligopeptide transport system. Twenty-seven Mcc HN-B genes were involved in the glycolysis/gluconeogenesis metabolic pathway. The metabolic pathways of glucose and pyruvate were intact, which may allow Mcc to generate energy under the hypoxic conditions.

3.3. Results of Comparative Genomic Analysis

3.3.1. Collinearity Analysis Results

We first confirmed the genomes of six strains (Mcc ATCC 27343, Mccp M1601, Mmc HN-A, MmmLC 95010, MmmSC PG1, and Ml PG50) in the Mm cluster as the target genomes (Table 2). The Mcc HN-B genome was used as the reference genome for collinearity analysis with target genomes.

	Mcc HN-B	Mcc ATCC 27343	Mccp M1601	Mmc HN-A	MmmLC 95010	MmmSC PG1	Ml PG50
Accession number	CP093215	CP000123.1	CP017125.1	CP093215	FQ377874.1	BX293980.2	CP002108.1
Isolation place	China	USA	China	China	France	Sweden	USA
Host	goat	goat	goat	goat	goat	cattle	cattle
Collection date	2022	2005	2016	2021	1995	2003	2010
Size (bp)	1,117,925	1,010,023	1,016,707	1,084,691	1,153,998	1,211,703	1,008,951
G + C (%)	23.79	23.8	23.67	23.76	23.81	24.0	23.8

Table 2. Basic information on each strain of the Mm cluster.

The collinearity results were statistically analyzed (Table S8). A plot of the parallel collinearity results was drawn based on these results (Figure 4). The relationships between Mcc ATCC 27343, Mccp M1601, and Mcc HN-B were the closest, followed by Ml PG50. MmmSC PG1 had the farthest genetic relationship with Mcc HN-B, followed by Mmc HN-A and MmmLC 95010. Many reverse-matching regions were found between the genomes of *Mycoplasma capricolum* and *Mycoplasma mycoides*. However, this phenomenon was not observed in Ml. Thus, the evolutionary distance between Ml and *Mycoplasma capricolum* genomes was closer than that between Ml and *Mycoplasma mycoides* genomes.



Figure 4. Parallel collinearity result plot. The upper axis represents the target species genome; the lower axis represents the Mcc HN-B genome. The red line indicates positive matching in the corresponding region; the blue line indicates reverse matching in the corresponding region.

3.3.2. Statistical Results of SNPs, InDels, and SVs

The number of SNPs, InDels, and SVs between Mcc HN-B and the target strains was determined and analyzed (Table 3). The figure of the variation types is shown (Figure S1).

Table 3. Statistical results of SNPs, InDels, and SVs between Mcc HN-B and target strains.

	SNPs	Insertions	Deletions	SVs
Mcc ATCC 27343	17,885	500	497	98
Mccp M1601	19,302	634	652	104
Mmc HN-A	32,999	806	841	123
MmmLC 95010	33,126	849	912	122
MmmSC PG1	25,573	613	683	141
Ml PG50	33,196	703	792	97

As shown in Table 3, the genomes of Mcc HN-B and Mcc ATCC 27343 had the lowest numbers of SNPs, insertions, and deletions. In addition, there was one more SV between Mcc HN-B and Mcc ATCC 27343 than between Mcc HN-B and Ml PG50. Therefore, Mcc

HN-B had the lowest genome variations when compared with Mcc ATCC 27343, followed by Mccp M1601. Combined with the picture of variation types (Figure S1), many similar inverted regions were found between *Mycoplasma capricolum* and *Mycoplasma mycoides*. These regions corresponded to the reverse-matching regions mentioned above. *Mycoplasma capricolum* and MI were more similar in genome structure than *Mycoplasma mycoides*.

3.4. Molecular Characterization of the Mm cluster

The Mm cluster is a group of mycoplasmas that belong to the genus *Mycoplasma*. It includes six important pathogenic mycoplasmas for ruminants: Mmc, MmmLC, MmmSC, Mcc, Mccp, and Ml. Therefore, finding the core genome, homologous gene family, and other molecular characteristics of the Mm cluster is of great significance for rapid clinical diagnosis, vaccine development, and disease treatment.

The six selected target genomes were different subspecies within the Mm cluster. Together with the Mcc HN-B genome, we performed gene family, core/pan-genome, phylogenetic tree, and ANI analyses to determine the molecular characteristics of the Mm cluster.

3.4.1. Gene Family Analysis and Core Genome Identification of the Mm Cluster

We first collected the results of gene family analysis for the seven mycoplasma strains (Table 4). A total of 580 orthologs of the Mm cluster and unique families in each strain were identified therein (Table S9, Figure 5).

Species	Total Genes	Gene in Families	Unclustered Genes	Families	Unique Families
Mcc HN-B	912	849	63	818	64
Mcc ATCC 27343	825	804	21	792	21
Mccp M1601	889	796	93	773	93
Mmc HN-A	848	826	22	802	22
MmmLC 95010	927	899	28	837	30
MmmSC PG1	1,017	919	98	777	115
Ml PG50	850	809	41	789	41

Table 4. Gene family analysis results of the Mm cluster.

Subsequently, we identified the core/pan-genomes with the characteristics of the Mm cluster using core/pan-genome analysis. The core genome contained 580 orthologs, which were key components for maintaining the basic vital activities and molecular characteristics of the Mm cluster. Among them, 18 orthologs encode 13 lipoproteins, 3 putative lipoproteins, and 2 putative prolipoproteins. These lipoproteins include one LPPA family lipoprotein, one LPPD family lipoprotein, and one MOLPALP family lipoprotein. They constitute the basic virulence of the Mm cluster, which provides a reference for further exploration of their virulence and pathogenic mechanisms.

The pan-genome comprised 1363 orthologs, which showed genetic diversity within the Mm cluster. All shared genes and specific genes of each strain were obtained (Table 5). This provides a basis for identifying the molecular characteristics and evolutionary distances between different subspecies within the Mm cluster.



Figure 5. Homologous gene family analysis of the Mm cluster. An orange ellipse represents a genome. The green circle in the middle represents the intersecting part of all the ellipses. The figure above the bracket shows the number of gene families, and the figure in the bracket means the total number of genes in the corresponding gene family.

Species	Total Genes	Shared Gene Number	Specific Gene Number
Mcc HN-B	912	847	65
Mcc ATCC 27343	825	804	21
Mccp M1601	889	796	93
Mmc HN-A	848	826	22
MmmLC 95010	927	895	32
MmmSC PG1	1,017	835	182
Ml PG50	850	809	41

Table 5. Specific genes to each strain within the Mm cluster.

3.4.2. Phylogenetic Tree and ANI Analysis

According to the orthologs' clustering results of the Mm cluster in Section 3.4.1, a phylogenetic tree was constructed based on the single-copy orthologs (Figure 6). We found that Mcc HN-B had the closest relative genetic distance to Mccp M1601, followed by Ml PG50. This was consistent with the analysis results of the genomic structural variations of each strain in Section 3.3.2. Moreover, Mcc HN-B had the farthest relative genetic distance from Mmc HN-A and MmmLC 95010. Compared with the other subspecies of *Mycoplasma mycoides*, the relative genetic distance between MmmSC PG1 and Ml PG50 was smaller. They had a closer genetic relationship, and their hosts were both cattle.



Figure 6. Phylogenetic tree based on the feature genes. The end of the branch represents different species. The node represents a common ancestor. The number on the branch shows its reliability. The branch length shows relative genetic distance within species.

Additionally, we discovered that Mcc HN-B had a smaller genetic distance from Mccp M1601 than Mcc ATCC 27343. Mccp M1601 was isolated from China in 2016, whereas Mcc ATCC 27343 was isolated from America in 2005. This result may be due to the geographic differences in Mcc and the consequences of the long-term evolution of these strains.

Finally, a heatmap based on ANI values was drawn to show the average base similarity of homologous fragments between the two strains (Figure 7). The ANI values among Mcc HN-B, Mccp M1601, and Mcc ATCC 27343 were all greater than 0.96. Comparatively, the ANI values were all lower than 0.95 between MI PG50 and the three strains mentioned above. The ANI values between Mmc HN-A and MmmLC 95010 were greater than 0.95. The ANI value of either of them with MmmSC PG1 was lower than 0.95, which could effectively distinguish them. Mcc HN-B showed the highest ANI value with Mccp M1601, which is in accordance with the phylogenetic tree analysis results.



Figure 7. Heatmap of ANI analysis.

3.5. Molecular Characterization of the Mycoplasma Capricolum Genome

Mcc HN-B, Mccp M1601, and Mcc ATCC 27343 belong to *Mycoplasma capricolum*. Orthologs of the three genomes listed in Table S9 were selected for comparison. A Venn diagram was constructed to show the results (Figure 8).



Figure 8. Venn diagram of orthologs in *Mycoplasma capricolum*. Note: Each circle represents orthologs of the strain.

Through comparative analysis, we identified 692 orthologs (Table S10) and 1058 orthologs in the core genome and pan-genomes of *Mycoplasma capricolum*, respectively. In the core genome, 36 orthologs encode 24 lipoproteins, 2 prolipoproteins, 8 putative lipoproteins, and 2 putative prolipoproteins. Apart from the common LPPA and LPPD family lipoproteins, one VmcF lipoprotein, one PARCEL family lipoprotein, and three MOLPALP family lipoproteins were predicted to exist in *Mycoplasma capricolum*.

Compared with the lipoproteins in the core genome of the Mm cluster, the 18 extra lipoproteins in *Mycoplasma capricolum*, such as VmcF lipoprotein and PARCEL family lipoprotein, may enhance the virulence of *Mycoplasma capricolum*. Additionally, they could be markers of *Mycoplasma capricolum*. Moreover, specific genes in the genomes of Mcc HN-B, Mcc ATCC 27343, and Mccp M1601 were also found, which added support for their specific identification.

Finally, Mcc HN-B shared 767 orthologs with Mcc ATCC 27343, which is 75 more orthologs than those in the core genome of *Mycoplasma capricolum*. These 75 orthologs were specific to Mcc and were of great significance for distinguishing between Mcc and Mccp. Meanwhile, it is of particular concern that Mcc HN-B shares 721 orthologs with Mccp M1601, which is 29 more orthologs than those in the core genome of *Mycoplasma capricolum*. As both Mcc HN-B and Mccp M1601 were isolated from China, these 29 orthologs may have regional characteristics in China.

3.6. Identification of Caprine-Host-Specific Orthologs

The hosts of Mcc, Mccp, and Mmc are all goats. To explore the caprine-host-specific orthologs of these strains in China, the orthologs of the three strains (Mcc HN-B, Mccp M1601, and Mmc HN-A) isolated from China were selected for comparison (Table S9). The results are shown in a Venn diagram (Figure 9).

The core and pan-genomes of the caprine-host strains contained 653 and 1133 orthologs, respectively (Figure 9). Caprine-host-specific orthologs were included in this core genome. Subsequently, MmmSC PG1 and MI PG50, both with bovine hosts, were selected for further screening (Figure 9). The core and pan-genomes of the bovine-host strains had 695 and 1011 orthologs, respectively. Finally, the caprine-specific core genome was compared with the bovine-specific core genome to obtain the number of caprine-hostspecific orthologs in China.



Figure 9. Venn diagram of caprine-host-specific orthologs. Note: Each circle represents orthologs of the strain.

Finally, 63 orthologs were identified as caprine-host-specific orthologs in China (Table S11). In total, 105 orthologs were identified as bovine-host-specific orthologs (Table S12). Among the caprine-host-specific orthologs, 11 orthologs encode seven lipoproteins, one prolipoprotein, and three putative lipoproteins. Five orthologs encode three membrane proteins and two putative membrane proteins. The other orthologs encode proteins, such as M17 peptidase, ATP-binding cassette transporter permease subunit, type I RM system subunit M, and type I RM system subunit R. We speculated that these genes might play an important role in host recognition, virulence regulation, substance transportation, and energy generation when goats are infected with Mcc, Mccp, or Mmc. These results provide a reference for the further study of the specific infection mechanisms of Mcc, Mccp, and Mmc in goats.

4. Conclusions

In this study, the Mcc HN-B strain was successfully isolated and identified from Hainan Island, China, for the first time. According to whole-genome sequencing analysis, we first identified three GIs in the Mcc HN-B genome. At the same time, we analyzed the type I RM and PTS systems of Mcc HN-B, revealing 14 potential virulence factors of the strain, which provides a theoretical basis for subsequent studies on their pathogenic mechanisms and vaccine development for Mcc.

By comparative genome analysis of Mcc HN-B and the other six members within the Mm cluster, the core genome of the Mm cluster and unique families of each strain were identified. The phylogenetic tree and heatmap of ANI analysis revealed evolutionary and affinity relationships between the members of the Mm cluster. This provides a new perspective for understanding the key biological functions and main molecular characteristics of all members within the Mm cluster. Moreover, the core genome provides a reference for virulence research and drug development against the Mm cluster. The specific genes identified in each strain can be applied as new markers to diagnose and distinguish members of the Mm cluster.

Finally, the core genome of *Mycoplasma capricolum*, represented by Chinese isolates and 63 caprine-host-specific orthologs, was identified. These findings could provide material support for the pathogenic mechanism research, rapid diagnosis, and drug development for *Mycoplasma capricolum*.

Supplementary Materials: The following supporting information can be downloaded at: https: //www.mdpi.com/article/10.3390/microorganisms10112298/s1, Figure S1: Variation type display diagram, (A) variation type display diagram of Mcc ATCC 2734 vs. Mcc HN-B, (B) variation type display diagram of Mccp M1601 vs. Mcc HN-B, (C) variation type display diagram of Ml PG50 vs. Mcc HN-B, (D) variation type display diagram of Mmc HN-A vs. Mcc HN-B, (E) variation type display diagram of MmmLC 95010 vs. Mcc HN-B, (F) variation type display diagram of MmmSC PG1 vs. Mcc HN-B; Table S1: Software information related to gene function analysis; Table S2: The potential CRISPRs in Mcc HN-B; Table S3: Proteins with transmembrane structure in Mcc HN-A; Table

potential CRISPRs in Mcc HN-B; Table S3: Proteins with transmembrane structure in Mcc HN-A; Table S4: Secreted proteins in Mcc HN-B; Table S5: Gene annotation statistics of Mcc HN-B; Table S6: Gene islands of Mcc HN-B; Table S7: VFDB database annotation results for Mcc HN-B; Table S8: Coverage statistics of collinearity alignment between target strains and Mcc HN-B; Table S9: Orthologs of each strain in the Mm cluster; Table S10: The core genome of *Mycoplasma capricolum*; Table S11: Number of caprine-host-specific orthologs; Table S12: Number of bovine-host-specific orthologs.

Author Contributions: Conceptualization, Z.Z., H.G. and F.W.; methodology, Z.Z., H.G. and M.H.; validation, H.L., Q.C. and W.Z.; formal analysis, Z.Z., W.Z. and X.L.; investigation, Z.Z, J.J. and M.H.; resources, C.M., L.D. and F.W.; data curation, Z.Z., Q.C. and S.C.; writing—original draft preparation, Z.Z., F.W. and H.G.; writing—review and editing, Z.Z., F.W., L.D. and H.G.; visualization, H.L., W.Z. and H.G.; supervision, F.W.; project administration, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Academician Innovation Platform Project of Hainan Province (No. YSPTZX202013); Hainan Innovation Center for Academician of Jin Ningyi; Hainan Innovation Center for Academician of Zhang Yong; China Agriculture Research System of MOF and MARA: CARS-38; and Special Funding Projects for Local Science and Technology Development Guided by the Central Committee (ZY2022HN09).

Institutional Review Board Statement: This experimental protocol was reviewed and approved by the Hainan University Institutional Animal Use and Care Committee under the ethical approval code HNUAUCC-2022-000147.

Informed Consent Statement: Informed consent was obtained from all animal owners involved in the study.

Data Availability Statement: All the analyzed datasets in the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Fischer, A.; Shapiro, B.; Muriuki, C.; Heller, M.; Schnee, C.; Bongcam-Rudloff, E.; Vilei, E.M.; Frey, J.; Jores, J. The origin of the '*Mycoplasma mycoides* cluster' coincides with domestication of ruminants. *PLoS ONE* **2012**, *7*, e36150. [CrossRef] [PubMed]
- 2. Gómez-Martín, A.; Amores, J.; Paterna, A.; De la Fe, C. Contagious agalactia due to *Mycoplasma* spp. in small dairy ruminants: Epidemiology and prospects for diagnosis and control. *Vet. J.* **2013**, *198*, 48–56. [CrossRef] [PubMed]
- Cottew, G.S.; Breard, A.; DaMassa, A.J.; Ernø, H.; Leach, R.H.; Lefevre, P.C.; Rodwell, A.W.; Smith, G.R. Taxonomy of the Mycoplasma mycoides cluster. Isr. J. Med. Sci. 1987, 23, 632–635. [PubMed]
- 4. Thiaucourt, F.; Bölske, G. Contagious caprine pleuropneumonia and other pulmonary mycoplasmoses of sheep and goats. *Rev. Sci. Tech.* **1996**, *15*, 1397–1414. [CrossRef]
- 5. Yatoo, M.L.; Parray, O.R.; Bashir, S.T.; Bhat, R.A.; Gopalakrishnan, A.; Karthik, K.; Dhama, K.; Singh, S.V. Contagious caprine pleuropneumonia—A comprehensive review. *Vet. Q.* **2019**, *39*, 1–25. [CrossRef]
- Westberg, J.; Persson, A.; Holmberg, A.; Goesmann, A.; Lundeberg, J.; Johansson, K.E.; Pettersson, B.; Uhlén, M. The genome sequence of *Mycoplasma mycoides* subsp. mycoides SC type strain PG1T, the causative agent of contagious bovine pleuropneumonia (CBPP). *Genome Res.* 2004, 14, 221–227. [CrossRef]
- 7. Bergonier, D.; Berthelot, X.; Poumarat, F. Contagious agalactia of small ruminants: Current knowledge concerning epidemiology, diagnosis and control. *Rev. Sci. Tech.* **1997**, *16*, 848–873. [CrossRef]
- 8. Rodriguez, J.L.; DaMassa, A.J.; Brooks, D.L. Caprine abortion following exposure to *Mycoplasma capricolum* subsp. *capricolum*. J. *Vet. Diagn. Investig.* **1996**, *8*, 492–494. [CrossRef]
- 9. Nicolas, M.M.; Stalis, I.H.; Clippinger, T.L.; Busch, M.; Nordhausen, R.; Maalouf, G.; Schrenzel, M.D. Systemic disease in Vaal rhebok (Pelea capreolus) caused by mycoplasmas in the mycoides cluster. *J. Clin. Microbiol.* **2005**, *43*, 1330–1340. [CrossRef]

- De la Fe, C.; Gutiérrez, A.; Poveda, J.B.; Assunção, P.; Ramírez, A.S.; Fabelo, F. First isolation of *Mycoplasma capricolum* subsp. *capricolum*, one of the causal agents of caprine contagious agalactia, on the island of Lanzarote (Spain). *Vet. J.* 2007, 173, 440–442. [CrossRef]
- Pinho, L.; Thompson, G.; Machado, M.; Carvalheira, J. Management practices associated with the bulk tank milk prevalence of *Mycoplasma* spp. in dairy herds in Northwestern Portugal. *Prev. Vet. Med.* 2013, 108, 21–27. [CrossRef] [PubMed]
- Rodriguez, J.L.; Ermel, R.W.; Kenny, T.P.; Brooks, D.L.; DaMassa, A.J. Polymerase chain reaction and restriction endonuclease digestion for selected members of the "*Mycoplasma mycoides* cluster" and *Mycoplasma putrefaciens*. J. Vet. Diagn. Investig. 1997, 9, 186–190. [CrossRef] [PubMed]
- 13. Maigre, L.; Citti, C.; Marenda, M.; Poumarat, F.; Tardy, F. Suppression-subtractive hybridization as a strategy to identify taxonspecific sequences within the *Mycoplasma mycoides* Cluster: Design and validation of an *M. capricolum* subsp. *capricolum*-specific PCR assay. *J. Clin. Microbiol.* **2008**, *46*, 1307–1316. [CrossRef]
- Seersholm, F.V.; Fischer, A.; Heller, M.; Jores, J.; Sachse, K.; Mourier, T.; Hansen, A.J. Draft Genome Sequence of the First Human Isolate of the Ruminant Pathogen *Mycoplasma capricolum* subsp. *capricolum*. *Genome Announc*. 2015, 3, e00583-15. [CrossRef] [PubMed]
- Calcutt, M.J.; Foecking, M.F. Comparative analysis of the *Mycoplasma capricolum* subsp. *capricolum* GM508D genome reveals subrogation of phase-vZariable contingency genes and a novel integrated genetic element. *Pathog. Dis.* 2015, 73, ftv041. [CrossRef] [PubMed]
- Tatay-Dualde, J.; Prats-van der Ham, M.; de la Fe, C.; Paterna, A.; Sánchez, A.; Corrales, J.C.; Contreras, A.; Tola, S.; Gómez-Martin, Á. Antimicrobial susceptibility and multilocus sequence typing of *Mycoplasma capricolum* subsp. *capricolum*. *PLoS ONE* 2017, 12, e0174700. [CrossRef]
- 17. Tatay-Dualde, J.; Prats-van der Ham, M.; de la Fe, C.; Paterna, A.; Sánchez, A.; Corrales, J.C.; Contreras, A.; Gómez-Martin, Á. Resistance mechanisms against quinolones in *Mycoplasma capricolum* subsp. *capricolum*. *Vet. J.* **2017**, 223, 1–4. [CrossRef]
- 18. Eden, P.A.; Schmidt, T.M.; Blakemore, R.P.; Pace, N.R. Phylogenetic analysis of *Aquaspirillum magnetotacticum* using polymerase chain reaction-amplified 16S rRNA-specific DNA. *Int. J. Syst. Bacteriol.* **1991**, *41*, 324–325. [CrossRef]
- Lin, Y.; Yuan, J.; Kolmogorov, M.; Shen, M.W.; Chaisson, M.; Pevzner, P.A. Assembly of long error-prone reads using de Bruijn graphs. Proc. Natl. Acad. Sci. USA 2016, 113, E8396–E8405. [CrossRef]
- 20. Chen, S.F.; Zhou, Y.Q.; Chen, Y.R.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018, 34, 884–890. [CrossRef]
- Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.D.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome. *PLoS ONE* 2014, 9, e112963. [CrossRef] [PubMed]
- Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016, 44, 6614–6624. [CrossRef] [PubMed]
- Lagesen, K.; Hallin, P.; Rødland, E.A.; Staerfeldt, H.H.; Rognes, T.; Ussery, D.W. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007, 35, 3100–3108. [CrossRef]
- Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016, 44, W54–W57. [CrossRef] [PubMed]
- Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 2013, 29, 2933–2935. [CrossRef]
 [PubMed]
- Bertelli, C.; Brinkman, F.S.L. Improved genomic island predictions with IslandPath DIMOB. *Bioinformatics* 2018, 34, 2161–2167. [CrossRef] [PubMed]
- Grissa, I.; Vergnaud, G.; Pourcel, C. CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 2007, 35, W52–W57. [CrossRef] [PubMed]
- 28. Haas, B. TransposonPSI: An Application of PSI-Blast to Mine (Retro-) Transposon ORF Homologies. Available online: http: //transposonpsi.sourceforge.net/ (accessed on 1 October 2022).
- 29. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinform. 2004, 4, 4.10.1-4.10.14.
- 30. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 1999, 27, 573–580. [CrossRef]
- 31. Fouts, D.E. Phage_Finder: Automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res.* **2006**, *34*, 5839–5851. [CrossRef]
- 32. Petersen, T.N.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* 2011, *8*, 785–786. [CrossRef] [PubMed]
- Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. J. Mol. Biol. 2001, 305, 567–580. [CrossRef] [PubMed]
- 34. Arnold, R.; Brandmaier, S.; Kleine, F.; Tischler, P.; Heinz, E.; Behrens, S.; Niinikoski, A.; Mewes, H.W.; Horn, M.; Rattei, T. Sequence-Based Prediction of Type III Secreted Proteins. *PLoS Pathog.* **2009**, *5*, e1000376. [CrossRef]
- Blin, K.; Wolf, T.; Chevrette, M.G.; Lu, X.; Schwalen, C.J.; Kautsar, S.A.; Suarez Duran, H.G.; de Los Santos, E.L.C.; Kim, H.U.; Nave, M.; et al. AntiSMASH 4.0-Improvements in Chemistry Rrediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* 2017, 45, W36–W41. [CrossRef] [PubMed]

- 36. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and Open Software for Comparing Large Genomes. *Genome Biol.* 2004, *5*, R12. [CrossRef] [PubMed]
- Goel, M.; Sun, H.; Jiao, W.B.; Schneeberger, K. SyRI: Finding Genomic Rearrangements and Local Sequence Differences from Whole-Genome Assemblies. *Genome Biol.* 2019, 20, 277. [CrossRef] [PubMed]
- Buchfink, B.; Reuter, K.; Drost, H.G. Sensitive Protein Alignments at Tree-of-Life Scale using DIAMOND. *Nat. Methods* 2021, 18, 366–368. [CrossRef]
- Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* 2003, 13, 2178–2189. [CrossRef]
- 40. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform*. 2004, *5*, 113. [CrossRef]
- Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* 2000, 17, 540–552. [CrossRef]
- 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. *Mol. Biol. Evol.* 2015, *32*, 268–274. [CrossRef] [PubMed]
- 43. Pritchard, L.; Glover, R.H.; Humphris, S.; Elphinstone, J.G.; Toth, I.K. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* **2016**, *8*, 12–24. [CrossRef]
- Qiu, X.B.; Shao, Y.M.; Miao, S.; Wang, L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol. Life Sci.* 2006, 63, 2560–2570. [CrossRef] [PubMed]
- Xie, Z.; Tang, H. ISEScan: Automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 2017, 33, 3340–3347. [CrossRef] [PubMed]
- Sirover, M.A. New insights into an old protein: The functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase. *Biochim. Biophys Acta* 1999, 1432, 159–184. [CrossRef]
- Hoelzle, L.E.; Hoelzle, K.; Helbling, M.; Aupperle, H.; Schoon, H.A.; Ritzmann, M.; Heinritzi, K.; Felder, K.M.; Wittenbrink, M.M. MSG1, a surface-localised protein of *Mycoplasma suis* is involved in the adhesion to erythrocytes. *Microbes Infect.* 2007, *9*, 466–474. [CrossRef] [PubMed]
- Goebel, W.; Chakraborty, T.; Kreft, J. Bacterial hemolysins as virulence factors. *Antonie Van Leeuwenhoek* 1988, 54, 453–463. [CrossRef] [PubMed]
- Mengaud, J.; Vicente, M.F.; Chenevert, J.; Pereira, J.M.; Geoffroy, C.; Gicquel-Sanzey, B.; Baquero, F.; Perez-Diaz, J.C.; Cossart, P. Expression in *Escherichia coli* and sequence analysis of the listeriolysin O determinant of Listeria monocytogenes. *Infect Immun.* 1988, 56, 766–772. [CrossRef]
- Alm, R.A.; Stroeher, U.H.; Manning, P.A. Extracellular proteins of Vibrio cholerae: Nucleotide sequence of the structural gene (hlyA) for the haemolysin of the haemolytic El Tor strain 017 and characterization of the hlyA mutation in the non-haemolytic classical strain 569B. *Mol. Microbiol.* 1988, 2, 481–488.
- 51. Li, Y.; Wang, R.; Sun, W.; Song, Z.; Bai, F.; Zheng, H.; Xin, J. Comparative genomics analysis of *Mycoplasma capricolum* subsp. *capripneumoniae* 87001. *Genomics* 2020, 112, 615–620. [CrossRef]
- 52. Balasubramanian, S.; Kannan, T.R.; Baseman, J.B. The surface-exposed carboxyl region of *Mycoplasma pneumoniae* elongation factor Tu interacts with fibronectin. *Infect. Immun.* **2008**, *76*, 3116–3123. [CrossRef] [PubMed]
- Pinto, P.M.; Chemale, G.; de Castro, L.A.; Costa, A.P.; Kich, J.D.; Vainstein, M.H.; Zaha, A.; Ferreira, H.B. Proteomic survey of the pathogenic *Mycoplasma hyopneumoniae* strain 7448 and identification of novel post-translationally modified and antigenic proteins. *Vet. Microbiol.* 2007, 121, 83–93. [CrossRef] [PubMed]
- 54. Dallo, S.F.; Kannan, T.R.; Blaylock, M.W.; Baseman, J.B. Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in *Mycoplasma pneumoniae*. *Mol. Microbiol.* **2002**, *46*, 1041–1051. [CrossRef] [PubMed]
- 55. Thomas, C.; Jacobs, E.; Dumke, R. Characterization of pyruvate dehydrogenase subunit B and enolase as plasminogen-binding proteins in *Mycoplasma pneumoniae*. *Microbiology* **2013**, *159*, 352–365. [CrossRef]
- 56. Hoelzle, L.E. Haemotrophic mycoplasmas: Recent advances in Mycoplasma suis. Vet. Microbiol. 2008, 130, 215–226. [CrossRef]
- 57. Esgleas, M.; Li, Y.; Hancock, M.A.; Harel, J.; Dubreuil, J.D.; Gottschalk, M. Isolation and characterization of alpha-enolase, a novel fibronectin-binding protein from *Streptococcus suis*. *Microbiology* **2008**, 154, 2668–2679. [CrossRef]
- 58. Bergmann, S.; Rohde, M.; Chhatwal, G.S.; Hammerschmidt, S. α-Enolase of *Streptococcus pneumoniae* is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol. Microbiol.* **2001**, *40*, 1273–1287. [CrossRef]
- 59. Allam, A.B.; Brown, M.B.; Reyes, L. Disruption of the S41 peptidase gene in *Mycoplasma mycoides* capri impacts proteome profile, H(₂)O(₂) production, and sensitivity to heat shock. *PLoS ONE* **2012**, *7*, e51345. [CrossRef]
- De Ste Croix, M.; Vacca, I.; Kwun, M.J.; Ralph, J.D.; Bentley, S.D.; Haigh, R.; Croucher, N.J.; Oggioni, M.R. Phase-variable methylation and epigenetic regulation by type I restriction-modification systems. *FEMS Microbiol. Rev.* 2017, 41, S3–S15. [CrossRef]
- 61. Doberenz, S.; Eckweiler, D.; Reichert, O.; Jensen, V.; Bunk, B.; Spröer, C.; Kordes, A.; Frangipani, E.; Luong, K.; Korlach, J.; et al. Identification of a *Pseudomonas aeruginosa* PAO1 DNA methyltransferase, its targets, and physiological roles. *mBio* 2017, *8*, e02312-16. [CrossRef]

- Nye, T.M.; Jacob, K.M.; Holley, E.K.; Nevarez, J.M.; Dawid, S.; Simmons, L.A.; Watson, M.E. DNA methylation from a Type I restriction modification system influences gene expression and virulence in *Streptococcus pyogenes*. *PLoS Pathog.* 2019, 15, e1007841. [CrossRef] [PubMed]
- 63. Kundig, W.; Ghosh, S.; Roseman, S. Phosphate Bound to Histidine in a Protein as an Intermediate in a Novel Phospho-Transferase System. *Proc. Natl. Acad. Sci. USA* **1964**, *52*, 1067–1074. [CrossRef]
- 64. Deutscher, J.; Francke, C.; Postma, P.W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* 2006, 70, 939–1031. [CrossRef] [PubMed]
- 65. Crigler, J.; Bannerman-Akwei, L.; Cole, A.E.; Eiteman, M.A.; Altman, E. Glucose can be transported and utilized in *Escherichia coli* by an altered or overproduced N-acetylglucosamine phosphotransferase system (PTS). *Microbiology* **2018**, *164*, 163–172. [CrossRef] [PubMed]
- 66. Voigt, C.; Bahl, H.; Fischer, R.J. Identification of PTS(Fru) as the major fructose uptake system of *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 7161–7172. [CrossRef]