

RESEARCH

Open Access



Unravelling the main genomic features of *Mycoplasma equirhinis*

Matthieu Martineau^{1,2,3†}, Chloé Ambroset^{3†}, Stéphanie Lefebvre³, Éléna Kokabi^{1,2}, Albertine Léon^{1,2} and Florence Tardy^{3,4*}

Abstract

Background *Mycoplasma* spp. are wall-less bacteria with small genomes (usually 0.5–1.5 Mb). Many *Mycoplasma* (*M.*) species are known to colonize the respiratory tract of both humans and livestock animals, where they act as primary pathogens or opportunists. *M. equirhinis* was described for the first time in 1975 in horses but has been poorly studied since, despite regular reports of around 14% prevalence in equine respiratory disorders. We recently showed that *M. equirhinis* is not a primary pathogen but could play a role in co-infections of the respiratory tract. This study was a set up to propose the first genomic characterization to better our understanding of the *M. equirhinis* species.

Results Four circularized genomes, two of which were generated here, were compared in terms of synteny, gene content, and specific features associated with virulence or genome plasticity. An additional 20 scaffold-level genomes were used to analyse intra-species diversity through a pangenome phylogenetic approach. The *M. equirhinis* species showed consistent genomic homogeneity, pointing to potential clonality of isolates despite their varied geographical origins (UK, Japan and various places in France). Three different classes of mobile genetic elements have been detected: insertion sequences related to the IS1634 family, a putative prophage related to *M. arthritidis* and integrative conjugative elements related to *M. arginini*. The core genome harbours the typical putative virulence-associated genes of mycoplasmas mainly involved in cytoadherence and immune escape.

Conclusion *M. equirhinis* is a highly syntenic, homogeneous species with a limited repertoire of mobile genetic elements and putative virulence genes.

Keywords Mycoplasma, Genome, Mobile genetic elements, Diversity, Virulence

[†]Matthieu Martineau and Chloé Ambroset contributed equally to this work.

*Correspondence:

Florence Tardy
florence.tardy@anses.fr

¹Research Department, LABÉO, Saint-Contest, Caen F-14000, France

²University of Caen Normandie, University of Rouen Normandie, INSERM, Normandie Univ, DYNAMICURE UMR 1311, Caen F-14000, France

³University of Lyon, Anses, VetAgro Sup, UMR Animal Mycoplasmosis, Lyon F-69007, France

⁴Present address: Anses, Ploufragan-Plouzané-Niort Laboratory—Mycoplasma, Bacteriology and Antimicrobial, Resistance Unit, Ploufragan F-22440, France

Background

The *Mycoplasma* genus comprises major recognized bacterial pathogens for both human and veterinary health [1]. *Mycoplasma* spp. in animals primarily colonize and infect mucosal areas of the respiratory and urogenital tract and the joints, while haemotrophic mycoplasmas infect red blood cells [1, 2]. However, concomitantly, other *Mycoplasma* species colonize the same physiological niches (oropharynx, respiratory and genital tracts) in both healthy and diseased individuals. These species are considered commensal, opportunistic, or of unknown status regarding their infectious potential.



The *Mycoplasma* genus species already counts over 120 species, and the number continues to rise as the science community reports new species in various hosts, but most of the time these new species do not attract attention from clinicians.

From a bacteriology standpoint, *Mycoplasma* spp. cumulate several atypical characteristics. Their cells are small and pleomorphic, as they have no cell wall. Their genomes are small, at 580 to 1359 kbp and 568 to 1037 coding sequences (CDS), have a low G+C content (23–40%), use UGA as a tryptophan codon instead of the universal stop, and have limited coding of complete metabolic pathways [1]. Consequently, mycoplasma growth is fastidious or even impossible for haemotrophic mycoplasmas in acellular media, and can only be achieved in sterol-enriched complex media. Finally, but crucially, *Mycoplasma* infection strategies do not rely on ‘classical’ virulence factors, with the exception of *M. pneumoniae* that secretes the CARDS (community-acquired respiratory distress syndrome toxin) exotoxin [3, 4].

The first description of the *M. equirhinis* species dates back to 1975 when Allam and Lemcke isolated four strains from the nasopharynx of thoroughbred horses that were serologically different from known species at that time and were able to hydrolyse arginine and rapidly form “film and spots” on agar plates [5]. As one of the four isolates came from a healthy horse, it immediately posed the question of whether *M. equirhinis* was a true pathogen. However, *M. equirhinis* failed to reproduce the disease in Welsh ponies experimentally infected by nasal route [6], and over the years, studies comparing the prevalence of *M. equirhinis* in diseased versus healthy animals failed to reach firm conclusions [7–9]. In a recent study including more than 1900 clinical observations, we showed that the prevalence of *M. equirhinis* remained steady whatever the clinical score of the sampled horse

but increased in the presence of other virus or bacteria such as *Streptococcus equi* subsp. *zooepidemicus* [9].

The present study was set up to learn more about *M. equirhinis* species through performing its genomic characterization. There is currently only scarce data on *M. equirhinis*, and there are only two genomes available in public databases under the denomination of *Metamycoplasma equirhinis*, from the recent still-controversial proposal to revise the taxonomy of the *Mycoplasma* genus based on phylogeny [10]. One is the genome of the type strain (NCTC 10148) originally isolated in 1975 by Allam and Lemcke in the UK [5], which is only available at contig level, i.e. 67 contigs but no circularized chromosome. However, while writing this article, another genome from an isolate named M432/72 was deposited on the NCBI website (GenBank: CP137845.1) and turns out to be the circular, closed version of the genome from the NCTC 10148 isolate, previously available only at the scaffold level (GenBank: GCA_006385185.1). Circularized genomes, in contrast to those of the scaffold level, provide the complete gene content of a genome as well as the genomic context and synteny of genes and hence are important as references for future assemblies. The other genome is recent (early 2023) and circularized, but unfortunately the metadata associated with the corresponding isolate (Myco-7) is unavailable, although we can reasonably posit that the isolate is related to research by Uchida-Fujii in 2021 [8]. Neither of these two genomes have yet been studied in detail.

Here we used comparative genomics data to characterize the species *M. equirhinis*. The circularized genomes of four isolates, i.e. Myco-7, M432/72 (NCTC 10148) and two other genomes generated here, were compared in terms of synteny, gene content, and specific features associated with virulence or genome plasticity. We further sequenced an additional 20 scaffold-level genomes to analyse intra-species diversity through pangenome analysis and a phylogenetic approach.

Table 1 Summary of the main characteristics of the four circularized *M. equirhinis* genomes

	<i>M. equirhinis</i> strains			
	F13465	F14026	M432/72	Myco-7
Isolation date	2020	2020	1972	NK
Genome size	691,532	738,996	706,186	719,195
G+C content (%)	28.1	28.2	28.0	28.1
ANI to Myco-7 (%)	98.82	98.81	98.89	/
Nr. of CDS	563	607	572	594
Nr. of tRNA	33	34	34	33
Nr. of rRNA (23S and 16S) set	3	4	3	3
Nr. of rRNA 5S	1	1	1	1
Nr. Of ICE	0	0	0	0
Nr. Of IS (remnant)	0	1	0	1

NK, Not Known; ANI, Average Nucleotide Identity; ICE, Integrative and Conjugative Element; IS, Insertion Sequence

Results

Overall comparison of the four circularized genomes

All four genomes showed a low G+C content (28.0–28.2%) and a high ANI (Average Nucleotide Identity) varying from 98.8% for F14026 to 98.9% for F13465 and M432/72 benchmarked against Myco-7 as reference (Table 1). The total number of predicted CDS (Coding Sequences) varied from 563 for strain F13465 to 572 for M432/72, 594 for strain Myco-7 and 607 for strain F14026. These CDS numbers are coherent with the genome sizes, which were slightly bigger for strain F14026 (738,996 bp versus 691,532 for strain F13465 and 706,186 for strain M432/72). All four annotated genomes contained the four genes implicated in the arginine dihydrolyase pathway, as already described in *M. hominis*

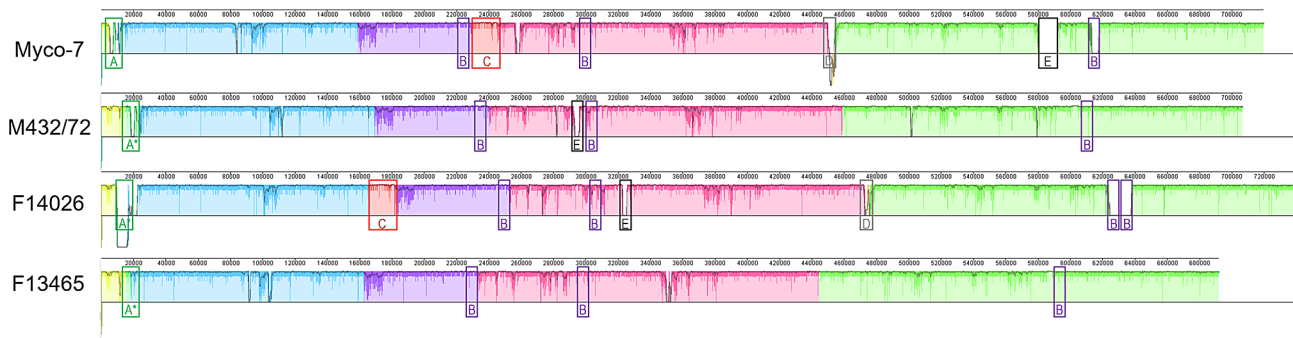


Fig. 1 Multiple genome alignment of F14026, F13465, M432/72 and Myco-7 *M. equirhinis* isolates. The alignment process generated a maximum of eight locally collinear blocks (LCB) representing homologous DNA regions shared between strains without sequence rearrangements. LCB order was set using the Myco-7 genome as a reference. Framed regions out of LCB correspond to A / A*, type 1 restriction system / CRISPR-Cas9; B, 16–23S sets; C, prophage; D, partial IS; E, hypothetical protein

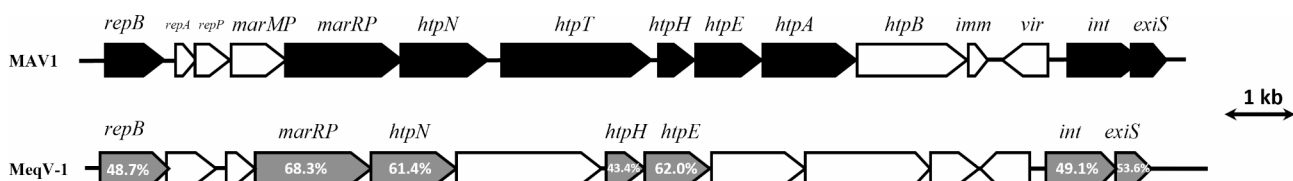


Fig. 2 Comparison of the MAV1 prophage (*M. arthritis*, isolate 158L3-1) and the MeqV-1 prophage-like sequence (*M. equirhinis*, isolate F14026). The two putative prophages share comparable overall ORF organization and orientation, but only seven proteins (in dark grey in the MeqV-1 phage) share >40% aa similarity. No homologues were identified for the other proteins. In MAV1, the different genes were assigned a function, as follows [18]: HtpN to HtpB, structural proteins of the MAV1 phage; RepB, replicative DNA helicase; RepA and P, replication initiators; MarMP, putative C5 methylase; MarRP, transcriptional regulator; Int, phage integrase; Exis, excisionase; Vir, protein to exclude superinfecting phage; Imm, phage repressor

species [11]. These four were genes *arcA* encoding arginine deiminase, *arcB* encoding ornithine carbamoyl-transferase, *arcC* encoding carbamate kinase, and a fourth gene encoding a dimethylarginine dimethylaminohydrolyase. This confirms the ability of *M. equirhinis* to hydrolyse arginine, as previously found in vitro [5].

Despite these homogeneous features, strain F14026 was different: it had four sets of 23S and 16S rRNA versus three for the three other strains (Table 1). PCR assays were run to rule out potential assembly errors. The number and size of the amplicons confirmed the presence of four 16S rRNA coding genes in isolate F14026 (Supplementary Fig. 1). Furthermore, strains F14026 and M432/72 had 34 (versus 33 for the two other strains) tRNA loci, due to a duplication of the gene coding the Aspartic acid tRNA, in a remote genomic location.

The four genomes were highly syntenic, with long locally collinear blocks (LCB) in the same order and orientation, suggesting few rearrangement events (Fig. 1). The few genomic regions that differed between the four genomes corresponded to: (i) a CRISPR-Cas9 element not found in Myco-7 that has a type 1 restriction-modification system at the corresponding location; (ii) a putative prophage and a vestigial IS (Insertion Sequence) detected in both Myco-7 and F14026 isolates, (iii) a length variable region of hypothetical proteins not found in F13465; (iv) the double copy of the 23S–16S set which

was specific to F14026. The two copies of the 23S–16S rRNAs were located in the same LCB of F14026 and only separated by a block of five CDS and a truncated copy of the *nusG* gene. This organization suggests a duplication event of the rRNA set as the upstream and downstream region of the corresponding 16–23S rRNA copy in isolate F13465, while the two copies in isolate F14026 were identical and corresponded to the five-CDS block on one side and the *nusG* gene on the other side. Interestingly, the 5S rRNA was a single copy in all four genomes, located at around 14 kb upstream of the third 16–23S rRNA set. A conserved prophage was detected in both Myco-7 and F14026 isolates (Fig. 1).

Detailed analysis of mobile genetic elements

The PHASTER analysis detected a 13.8-kb putative prophage (designated MEqV-1) in both the F14026 and Myco-7 genomes. This putative prophage displays a gene content and organization similar to those of the *M. arthritis* MAV-1 prophage group as described by Citti et al. [12]. MEqV-1 comprised 14 ORFs in a similar orientation to the 15 ORFs in MAV-1 (Fig. 2). However, of the 15 proteins from the MAV1 phage, only seven were found in F14026 with >40% amino acid identity (Fig. 2). Of the *rep* genes, only *repB*, encoding a DNA helicase, was detected in the putative prophage of strain F14026 while homologues of *repA*, encoding a protein that binds

DNA to initiate replication, and *repP*, encoding a suspected replication initiator, were missing. Similarly, for the *htp* genes involved in phage structure, only *htpN*, *H* and *E* were conserved between MAV1 and MeqV-1. The integrase (Int) and excisionase (Exis) as well as the genes encoded by the transcriptional regulator marRP were similar between MEqV-1 and MAV-1. The MEqV-1 sequence is conserved between the two *M. equirhinis* isolates (99.5% of nucleotide identity). It spans an entire LCB inserted in different places in the two genomes (Fig. 1), suggesting independent insertion events. No prophages were detected in the other scaffold-level genomes.

A gene encoding an IS transposase was detected in both F14026 and Myco-7 (but not in F13465 nor in M432/72). The two transposases are 627 amino acid (aa) long and have a pairwise aa identity of 98.7%. Their first BLASTp hit in ISFinder is an ISSeq4 transposase (IS1634 family) from *Streptococcus equi* of 573 aa (amino acid) with only 36% identity. In contrast, using the NCBI interface, the first BLASTp hit, with default parameters, is an IS1634-family transposase from a bacterium of the *Acholeplasma* order with 97% identity over 559 aa, and the first hit within the *Mycoplasma* genus is an IS1634-family transposase from *M. cottewii*, a commensal species of the ear canal in goats, with 42% identity on 628 aa. The conserved genomic organization around the *M. equirhinis* transposases in both the F14026 and Myco-7 genomes suggests that it is no longer mobile anymore and could be considered vestigial (Fig. 1). tBLASTn found one transposase similar to that of Myco-7 and F14026 in four of the draft genomes (isolates F13948, F14029, F14037 and F14358) with 100% query coverage and >99% identity.

The search for *Mycoplasma* Integrative and Conjugative Elements (MICE) of the *M. hominis* or *M. arginini* type in the circularized genomes only managed to find one copy of an ICE-Ho I-like CDS1 gene in F14026 and M432/72 (Supplementary Table 5). In both strains, this CDS1 was located in close proximity to the second copy of the gene coding Asp tRNA, suggesting a potential concomitant transfer. This CDS1-like gene was found on another nine isolates in the scaffold-level genomes (Supplementary Table 5). Furthermore, isolate F13947 harboured homologues of the four-CDS backbone of the *M. hominis* ICE-Ho I, albeit in four different contigs. Three backbone genes similar to the MICE of *M. arginini* (NGR_2017), i.e. CDS5, CDS17 and CDS22, and an additional *dnaG* gene characteristic of *M. arginini* MICE, were found on five draft genomes (F14022, F14029, F14346, F14358 and F14365). For four scaffold-level genomes, the ICE-related CDS were located on the same contig (Supplementary Table 5). Interestingly, the *M. equirhinis* MICE (MICE-Eql) in these isolates possess a variable region (from 0 to 4807 bp) containing zero to four ORFs located between CDS19 and 22 (Fig. 3).

Search for virulence genes or genes implicated in host-pathogen interaction

The nonspecific screening on the four circularized genomes highlighted only two potential virulence-associated genes: *tuf* coding an elongation factor Tu, and *hlya* coding a hemolysin. On top of its canonical role, the Tuf protein has also been shown to contribute to virulence through adhesion to host extracellular matrix components, dissemination, and evasion of the complement system [13]. However, as it is one of the most abundant

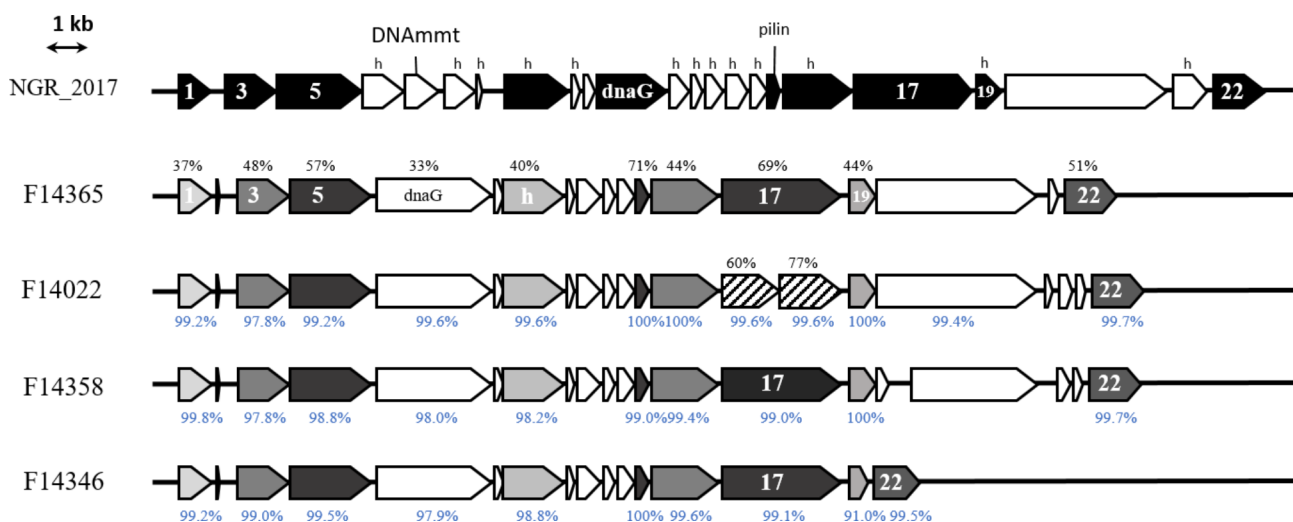


Fig. 3 Structural organization of MICEs detected in some *M. equirhinis* genomes, and comparison to the closest previously described MICE in *M. arginini* isolate NGR_2017. CDS are represented by arrows with their numbering inside. Hypothetical proteins with no associated function are marked 'h'. DNAmmt: DNA methyltransferase. Percentages in black correspond to the aa homology between *M. equirhinis* isolate F14365 and the corresponding proteins in *M. arginini* NGR_2017. Percentages in blue correspond to the aa homology between *M. equirhinis* isolate F14365 and the corresponding proteins in the three other *M. equirhinis* isolates. Pseudogenes are hatched

proteins in bacteria, its detection here does not mean it is effectively associated with virulence. In the circularized *M. equirhinis* genomes, the HlyA homologue was a 244-aa protein for isolates F14026, F13465 and M432/72. The HlyA protein was 100% identical between F13465 and M432/72 and showed 99.2%-aa identity with F14026, whereas in Myco-7 the protein (corresponding gene JMP7_3850) was truncated (207-aa). The first BLASTp hit of the conserved HlyA protein is a TlyA-family RNA methyltransferase from *M. arthritidis* (query coverage: 100%, percent identity: 68%), while the first hit using Phyre2 is indeed a putative haemolysin from *Streptococcus thermophilus* [14] with 99% query coverage despite only 35% identity for F14026, F13465 and M432/72 and 38% identity for Myco-7.

Table 2 summarizes the putative virulence-associated proteins of *M. hominis* found in the two French circularized genomes of *M. equirhinis*. These putative virulence-associated proteins include the MHO_0730 and MHO_0660 nucleases that could play a role in escaping NETs or METs, and several other proteins involved in cytoadherence and one copy of MIB-MIP system. All the corresponding genes, as well as those encoding HlyA and Tuf, were found in the core genome, suggesting their wide distribution among isolates.

Diversity of the species

The pangenome of the 24 *M. equirhinis* strains (22 French isolates in this study and two public sequences, namely Myco-7 from Japan and M432/72 from Great Britain) comprised 508 core genes amounting to 582,641 nucleotides, and 278 accessory genes. A phylogenetic tree inferred from the core-genome alignment using the Maximum of Likelihood (ML) method counted 16,374 parsimony-informative sites, resulting in 12,581 distinct patterns. In the resulting mid-rooted tree with no outgroups (Fig. 4), *M. equirhinis* isolates were split into

two different branches with very strong bootstrap values. Despite their different geographic origins and dates of isolation, strains Myco-7 and M432/72 appeared to be closely related to French isolates. We were unable to identify a specific tree structure based on geographic origin, time of sampling, or horse characteristics. Two strains, i.e. F13414 and F13415, shared an identical core-genome composition despite originating from two remote regions (the Pays de la Loire in northwest France and Bourgogne-Franche-Comté in central-eastern France) and having no epidemiological links.

Discussion

To our knowledge, this study is the first to propose an in-depth genomic characterization of *M. equirhinis*, an opportunistic species frequently isolated from horses with respiratory disorders [9]. Two complete genome sequences from two *M. equirhinis* strains isolated in France were generated using a hybrid assembly approach, and then compared against the Myco-7 and M432/72 (NCTC 10148) genomes available in public databases. One outstanding feature is the overall high degree of sequence similarity between the four strains (>98.8% shared identity for each isolate compared to Myco-7, Table 1) and their conserved synteny. This contrasts with the picture for other mycoplasma species involved in bovine respiratory infections, such as *M. bovis* or *M. mycoides* subsp. *mycoides* [15, 16], but is comparable to the picture for *M. pneumoniae* where there is over 99% identity between genomes [17]. These results suggest clonality of the four isolates despite their different geographic origins. We identified only one large region (13.8 kbp) of genomic structural variation corresponding to the insertion of a putative prophage in both strain Myco-7 and strain F14026. This prophage is related to the prophage described in *M. arthritidis* [18]. Another specific 11.3 kbp region was identified in strain Myco-7

Table 2 Putative virulence and cytoadherence proteins predicted in *M. equirhinis* F14026 and F13465 genomes in comparison to *M. hominis* strain ATCC 23114

	Function	Genes	Protein accession number	Protein homology in F14026		Protein homology in F13465		Found in the core genome
				Query cover (%)	ID (%)	Query cover (%)	ID (%)	
<i>M. hominis</i> ATCC 23114	Nucleases	MHO_0730	CAX37207	97.8	41.3	97.8	41.0	NO
		MHO_0660	CAX37200	71.6	44.6	71.6	45.0	YES
	MIB-MIP	MHO_3190	CAX37454	99.7	46.4	99.7	46.1	YES
		MHO_3200	CAX37455	99.4	45.9	99.4	45.9	YES
	Proteins potentially involved in cytoadherence	<i>P60</i>	CAX37484	99.6	45.9	99.6	45.9	YES
		<i>P80</i>	CAX37485	99.4	55.2	99.4	55.0	YES
		<i>P120</i>	CAX37515	100	49.6	100	49.6	YES
		<i>oppA</i>	CAX37285	99.4	42.9	99.4	42.5	YES

Abbreviations aa: amino acid; ANI: Average Nucleotide Identity; CARDS: community-acquired respiratory distress syndrome toxin; CDS: CoDing Sequence; ICE: Integrative conjugative element; IS: Insertion Sequence; kbp: kilo base pair; LCB: Locally Collinear Block; Mb: Megabits; MICE: Mycoplasma integrative and Conjugative Elements; ML: maximum likelihood; OD: Optical Density; ONT: Oxford Nanopore Technology; ORF: Open Reading Frame; PPLO: Pleuro-Pneumonia-Like-Organism; subsp.: subspecies; VFDB: Virulence Factor Database

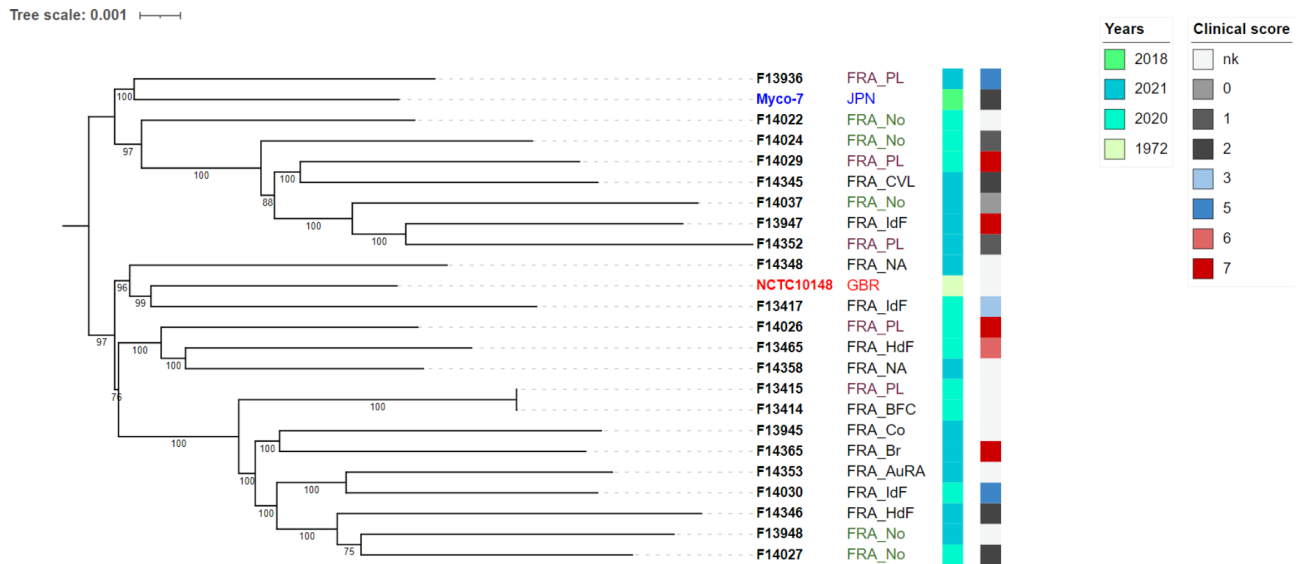


Fig. 4 Phylogenetic tree inferred from core-genome alignment of 24 *M. equirhinis* isolates. The core genome contained 508 genes. Statistical supports for branches were assessed using ultrafast bootstrap (1000 replicates). The phylogeny is midpoint-rooted and two main clusters can be distinguished. The branch lengths are drawn to scale and indicate the number of nucleotide substitutions per site. On the right of the tree are indicated the geographical origin of the isolates, their year of isolation and the associated cumulative score using the scoring system developed in Martineau et al. 2023 [9]. FRA, FRANCE; JPN, JAPAN; GBR, Great Britain. In France, the administrative region are abbreviated as follows: PL, Pays de la Loire; No, Normandie; CVL, Centre-Val de Loire; IdF, Île-de-France; NA, Nouvelle-Aquitaine; BFC, Bourgogne-Franche-Comté; Co, Corse; Br, Bretagne; AuRA, Auvergne-Rhône-Alpes; HdF, Hauts-de-France. In the tree, the isolate named NCTC10148 is the same than M432/72

but was not associated with any function (green box in Fig. 1).

Another outstanding feature was the fact that one of the circularized genomes has four 16S rRNA copies whereas the other three have only three. This characteristic was unique in a panel of 90 PCR-screened strains and could correspond to a duplication event. An IS transposase related to the IS1634 family was detected in the genomes of Myco-7 and F14026 at roughly the same genomic location, suggesting a non-mobile vestigial IS that was consequently not associated with genome rearrangements.

The set of analysed genomes was enlarged to 24 by including 20 additional scaffold-level genomes generated as part of this study. The 24 genomes share a core genome of 508 genes, representing 90%, 89%, 86% and 84% of the gene numbers of the four circularized genomes (F13465, M432/72, Myco-7 and F14026), respectively. Once again, this points to a high degree of similarity, with a reduced pool of accessory genes, in stark contrast to the picture in *M. hominis*, the closest phylogenetic neighbour, where the core genome represents only 55% of all genes [19]. The homogeneity of the species was confirmed by inferring a phylogeny from the core-genome alignment. Branches were overall short with a maximum of 25 substitutions per site between the furthest isolates. Although the isolates were clustered into two subgroups due to midpoint rooting, the ML tree showed no structuration linked with the clinical history or geographic origin of

the isolates. The isolates from Great Britain and Japan did not cluster separately, confirming this homogeneity could be a characteristic of the species, in agreement with the ANI > 98.8% between the four circularized genomes now available.

In the extended set of genomes, we also detected other MICE-type mobile genetic elements. A unique CDS, similar to the CDS1 of MICE-Ho I, was detected in 10 isolates, mainly in dissimilar gene contexts. Furthermore a MICE similar to that of *M. arginini* [20] was identified in five strains, with a relatively significantly conserved gene backbone. While the *M. hominis*-like CDS1 can easily be considered a vestige from a common ancestor of the two (phylogenetically very close) species, the origin of the *M. arginini*-like MICE raises questions. Figure 3 shows some differences in CDS organization between the four MICE that suggest that these elements are under strong evolutionary pressure but could still be active, as their DDE transposase and the genes associated with the conjugative process are conserved.

The IS transposase was detected in six genomes with identical adjacent genes, confirming loss of mobility and vertical inheritance of these elements. In the set of 24 isolates, the presence of the MEqV-1 prophage was limited to isolates F14026 and Myco-7, suggesting a low frequency of viral infections. The role of the detected CRISPR-Cas immune systems against this particular prophage invasion remains to be clarified [21].

All these mobile genetic elements are regularly identified within the *Mycoplasma* genus and are described as major contributors to genomic plasticity [12]. In *M. equirhina*, they were detected at low frequencies (25% of isolates harboured IS elements, 8% harboured putative prophages, and 25% harboured MICEs) compared to other species in the genus such as *M. bovis* and *M. hominis* [22, 23]. Even though these frequencies may be underestimated because most of the genomes were not circularized, this still stands as further evidence for low plasticity in *M. equirhina* genomes.

In silico analysis of putative virulence genes revealed that *M. equirhina* is equipped with cytoadherence and immune escape systems that are regularly reported in other species of the genus *Mycoplasma*. The repertoire of putative virulence-associated genes is very similar, albeit less extended, to that of *M. hominis*, confirming that *M. equirhina*, like *M. hominis*, could be opportunistic. All these genes were found in the core genome of *M. equirhina*, which again suggests homogeneous virulence. However, experimental tests (in vitro, ex vivo or in vivo) would be required to firmly conclude on the role of the associated proteins, especially in a context of co-infection. Indeed, any modulation of the immune response by *M. equirhina* could be favourable to colonization by other pathogens such as *S. zooepidemicus*, with which there is a known positive association [9]. If *M. equirhina* shares a similarly opportunistic nature to *M. hominis* despite the different host species and body niches, then it would be important to set up diagnostic tools able to quantify *M. equirhina* load at the site of infection, as already done for *M. hominis* [24].

Conclusion

In-depth analysis of *M. equirhina* genomes reveals a highly homogenous species, with only a few mobile genetic elements that potentially come from different origins but are not significantly associated to genome rearrangements. In terms of virulence potential, the core genome of the species harbours homologues of *M. hominis* genes that are mostly involved in cytoadherence or immune modulation. This genetic equipment is compatible with an opportunistic behaviour within its host. Additional sequencing from worldwide collected isolates could further these conclusions. Given this overall clonality of the *M. equirhina* isolates, it would be of interest, from a clinical perspective to develop tools to quantify the *M. equirhina* load at the site of infection in order to confirm whether clinical signs are associated to high loads.

Methods

Isolates, culture conditions, DNA extraction

For this work, we selected 22 *M. equirhina* isolates originating from a previous study [9] (Supplementary Table 1) that had been isolated from the lower respiratory tract of

horses from different geographical origins in France and presenting different associated clinical signs. For genome sequencing, a 48 h-culture broth of each isolate was filtered using a 0.22- μ m filter and then seeded onto PPLO agar plates (Indicia Production, St Genis L'Argentière). The plates were then incubated for 2 to 4 days at 37 °C and 5% CO₂, and one clone per strain was picked at random with a micropipette tip, and further subcultured. Genomic DNA was extracted from 2 to 5 culture tubes of 2 mL each, 48-h cultures of the clones using a commercial kit from Lucigen (Epicentre). The quality (OD260/280 \geq 1.8 and OD260/230 \geq 2) and concentration of the extracted DNA (at 6 μ g at 30–500 ng/ μ L per strain for Oxford Nanopore Technology (ONT) DNA sequencing and 100 ng at 5 ng/ μ L per strain for Illumina sequencing) were checked using a Nanodrop spectrophotometer (Thermo Fisher) and a Qubit fluorimeter (DNAds BR kit, Thermo Fisher), respectively.

Genome sequencing, assembly, and annotation

The 22 strains were sequenced using Illumina technology (Nextera XT DNA library preparation kit and a NovaSeq sequencer), and strains F13465 and F14026 were also sequenced using Oxford Nanopore technology (ONT) recommendations and kits on a MinION Flow Cell (R 10.4). Both Illumina and ONT sequencing were outsourced (Supplementary Table 2). Quality control of raw data (FASTQ files for both short and long reads), genome assemblies (draft genomes for short reads and circularized genomes for hybrid assemblies) and annotations were performed according to previously published protocols and parameters [15]. In brief, Illumina reads were trimmed using Trimmomatic (v.0.39) [25] and their quality was assessed using fastqc (v.0.11.5) [26]. Selected reads had a Phred score \geq Q25 and sizes comprised between 50 and 150 bp. Quality of the ONT long reads was checked using Longqc 1.2b [27]. Long reads were then trimmed using Filtlong 0.2.0 [28] using Illumina paired-end reads as external reference and with quality and read length set to Q14 and 30kbp, respectively. Assembly was done using Spades 3.14.1 [29, 30] for draft assemblies and Unicycler 0.4.8 [31, 32] for hybrid assemblies. Annotation was done using Prokka 1.14.6 [33, 34]. Assembly statistics and genome completeness were computed using QUAST 5.2.0 [35, 36] and Busco score; which was computed using Busco 5.5.0 [37, 38]. All these bioinformatics tools and pipelines were run using the IFB Core cluster resource [39].

Comparative genomic analysis of circularized genomes

The genomes of strain Myco-7 (GCF_030295865.1) and strain M432/72 (GenBank: CP137845.1) were included for comparison purposes together with the circularized genomes of strains F14026 and F13465. Pairwise

comparison of ANI was performed using fastANI v1.3 at the default parameter settings [40]. The synteny of the four circularized genomes (with Myco-7 taken as reference) was compared using ProgressiveMauve on Mauve (version 20150226) with the following parameters: 15 match seed weight and min LCB weight set to 10,000 [41, 42]. Genes found in regions differing between two genomes were identified using BLASTp on a ref-seq protein database [43].

PCR amplification for screening 16S rRNA copy numbers

Four PCR assays were designed and validated to check the copy numbers of the gene coding for 16S rRNA in the 22 isolates. One primer was located at the 5' terminal of the 16S rRNA gene and was designed using Primer3 as available under Geneious (Geneious Prime; Biomatters, New Zealand, v2023.2.1) on a multiple alignment of the seven 16S copies of isolates F13465 and F14026 (named 16S primer). The four other primers were designed to target genes upstream or downstream the 16S copies, using the consensus sequence of F13465 and F14026 alignments for primers 1 to 3 and only the sequence of F14026 for primer 4.

The four PCR assays were run in a 25 μ L PCR volume with Premix Ex TAQ (Takara), i.e. 'Condition I' (Supplementary Table 3) on the 22 DNAs from the sequenced isolates. The PCR assay targeting the 4th copy was also conducted in a 25 μ L PCR volume with Go Taq (Promega), i.e. 'Condition II', on an enlarged set of 90 DNAs including the 22 that were sequenced here. The sizes of the amplicons were verified by capillary electrophoresis using a QIAxcel Advanced system (Qiagen).

In silico analyses to detect mobile genetic elements and potential virulence factors

These analyses were run in both circularized genomes and scaffold-level genomes generated here.

The search for potential virulence genes was first carried out using the generic VFanalyser pipeline and the VFDB database [44, 45]. Then, a selection of mycoplasma-specific virulence-associated proteins were searched by BLASTp against a custom protein database encompassing all the sequenced genomes of *M. equirehinis* that was built using the Geneious environment. The query included (i) mycoplasma immunoglobulin binding protein and the associated mycoplasma immunoglobulin protease (i.e. 'MIB-MIP system'), proteins potentially involved in cytoadherence and nucleases described in the very close species *M. hominis* [11, 46, 47], (ii) variable surface proteins on the model of *M. bovis* [48, 49], (iii) proteins involved in the biosynthesis of capsular polysaccharides in the *M. agalactiae* model [50]; and (iv) the CARDS toxins from *M. pneumoniae* (See Supplementary Table 4 for details on the accession numbers of the different queries). Only hits with $\geq 60\%$ cover and $\geq 40\%$ pairwise

identity were considered significant and are presented in the **results** section.

We investigated the presence of insertion sequence elements (IS) in the four circularized genomes using ISEScan (1.7.2.3) on the Galaxy Europe platform with default parameter settings except for the option to remove incomplete IS [51, 52]. In case of IS detection by ISEScan, the potential transposase protein was identified using the BLASTp function on the ISFinder database [53, 54]. Any IS transposases identified in the circularized genomes were further searched by tBLASTn in the other genomes.

For the detection of integrative and conjugative elements (ICE), the four CDS corresponding to the ICE backbone (i.e. the two extremities CDS1 and CDS 22, as well as CDS5 and CDS17 that encode homologues of TraE and TraG, respectively, two transmembrane proteins with conserved ATPase domains involved in the conjugative machinery) of the two variant copies described in *M. hominis*, i.e. ICEHo-I and -II, and the ICE of *M. arginini* [20, 55, 56], were searched by BLASTn against a custom nucleotide database encompassing all the sequenced genomes of *M. equirehinis* built using the Geneious environment (Supplementary Table 5). The presence of the *M. arginini* ICE was further confirmed by searching *dnaG*. These two species were chosen based on the results of the Prokka annotation.

We used the PHASTER web server [57–59] to explore the presence of a prophage sequence. The analysis was then refined by determining the percent identity of individual proteins shared with the MAV1 prophage from *M. arthritidis*, using Clustal Omega (1.2.2) run in Geneious. Phyre2 was further used to gain deeper insight into the putative function of certain proteins with no identity [60, 61].

Phylogenetic tree construction

The core genome of the population (i.e. the 22 strains sequenced in this study plus the two complete genomes available at NCBI) was calculated using the Panaroo pipeline 1.2.10 [62, 63] on the IFB Core cluster [39]. The resulting core-genome alignment was used to construct a maximum likelihood (ML) phylogeny in IQ-TREE v2.0.3 [64]. General time reversible nucleotide substitution model, gamma distribution, and proportion of invariable sites (GTR+G+I) were selected as the best-fit model parameters for our data by IQ-TREE. Statistical support for node of the ML phylogeny was assessed using ultrafast bootstrap approach of 1000 replicates. Visualisation of the core-genome phylogenetic tree was done with iTOL [65, 66] and the tree was midpoint rooted.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10789-y>.

Supplementary Material 1
 Supplementary Material 2
 Supplementary Material 3
 Supplementary Material 4
 Supplementary Material 5
 Supplementary Material 6
 Supplementary Material 7

Acknowledgements

We thank Sophie CASTAGNET and Agnès TRICOT for their help isolating and preserving the strains used in this study. We also thank Dr. Chervin HASSEL for his help with the genome deposition process.

Author contributions

M. M. and C.A.: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – Original draft. S. L. and E.K.: Data curation, Formal analysis. A. L.: Conceptualization, Funding acquisition, Supervision, Visualization. F. T.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – Original draft, Writing – Review and Editing. All authors are aware of the content of the manuscript.

Funding

This work received financial support from the IFCE (Institut Français du Cheval et de l'Équitation, grant number CS-2020-2023-022-MYCOPAB), the Fonds Eperon (grant number N15-2020), and the GIS (Groupement d'Intérêt Scientifique) CENTAURE (grant number GIS20-SEP-01).

Data availability

Genomic data that support the findings of this study have been deposited under the BioProject PRJNA1045244, accessible at <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1045244>. Details on the associated accession numbers are provided in Supplementary Table 2.

Declarations

Ethics approval and consent to participate

No ethical approval required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 April 2024 / Accepted: 10 September 2024

Published online: 20 September 2024

References

- Brown DR, May M, Bradbury JM, Balish MF. *Bergey's Manual of Systematics of Archaea and Bacteria*. 1st edition. Wiley; 2018.
- Messick JB. Hemotrophic mycoplasmas (hemoplasmas): a review and new insights into pathogenic potential. *Vet Clin Pathol*. 2004;33:2–13.
- Kannan TR, Baseman JB. ADP-ribosylating and vacuolating cytotoxin of *Mycoplasma pneumoniae* represents unique virulence determinant among bacterial pathogens. *Proc Natl Acad Sci U S A*. 2006;103:6724–9.
- Yiwen C, Yueyue W, Lianmei Q, Cuiming Z, Xiaoxing Y. Infection strategies of mycoplasmas: unraveling the panoply of virulence factors. *Virulence*. 2021;12:788–817.
- Allam NM, Lemcke RM. Mycoplasmas isolated from the respiratory tract of horses. *J Hyg Camb*. 1975;74:29.
- Hooker JM, Butler M. The development of complement-fixing antibody in horses after infection with *Mycoplasma equirhinis*. *J Comp Pathol*. 1977;87:281–6.
- Wood JLN, Newton JR, Chanter N, Mumford JA. Association between respiratory disease and bacterial and viral infections in British racehorses. *J Clin Microbiol*. 2005;43:120–6.
- Uchida-Fujii E, Kinoshita Y, Niwa H, Maeda T, Nukada T, Ueno T. High prevalence of *Mycoplasma equirhinis* in thoroughbred horses with respiratory symptoms in autumn 2018. *J Vet Med Sci*. 2021;83:1907–12.
- Martineau M, Kokabi E, Taiebi A, Lefebvre S, Pradier S, Jay M, et al. Epidemiology and pathogenicity of *M. equirhinis* in equine respiratory disorders. *Vet Microbiol*. 2023;287:109926.
- Gupta RS, Sawnani S, Adeolu M, Alnajjar S, Oren A. Phylogenetic framework for the phylum Tenericutes based on genome sequence data: proposal for the creation of a new order *Mycoplasmoidales* ord. nov., containing two new families *Mycoplasmoidaceae* fam. nov. and *Metamycoplasmataceae* fam. nov. harbouring *Eperythrozoon*, *Ureaplasma* and five novel genera. *Antonie van Leeuwenhoek*. 2018;111:1583–630.
- Pereyre S, Sirand-Pugnet P, Beven L, Charron A, Renaudin H, Barré A, et al. Life on arginine for *Mycoplasma hominis*: clues from its minimal genome and comparison with other human urogenital mycoplasmas. *PLoS Genet*. 2009;5:e1000677.
- Citti C, Baranowski E, Dordet-Frisoni E, Faucher M, Nouvel L-X. Genomic islands in Mycoplasmas. *Genes*. 2020;11:836.
- Harvey KL, Jarocki VM, Charles IG, Djordjevic SP. The diverse functional roles of elongation factor Tu (Ef-Tu) in microbial pathogenesis. *Front Microbiol*. 2019;10:2351.
- Putative. hemolysin from *Streptococcus thermophilus*. https://www.wwpdb.org/pdb?id=pdb_00003hp7
- Ambroset C, Peticca A, Tricot A, Tardy F. Genomic features of *Mycoplasma bovis* subtypes currently circulating in France. *BMC Genomics*. 2022;23:603.
- Dupuy V, Manso-Silván L, Barbe V, Thebault P, Dordet-Frisoni E, Citti C, et al. Evolutionary history of contagious bovine pleuropneumonia using next generation sequencing of *Mycoplasma mycoides* subsp. *mycoides* small colony. *PLoS ONE*. 2012;7:e46821.
- Xiao L, Ptacek T, Osborne JD, Crabb DM, Simmons WL, Lefkowitz EJ, et al. Comparative genome analysis of *Mycoplasma pneumoniae*. *BMC Genomics*. 2015;16:610.
- Voelker LL, Dybvig K. Characterization of the lysogenic bacteriophage MAV1 from *Mycoplasma arthritidis*. *J Bacteriol*. 1998;180:5928–31.
- Jironkin A, Brown RJ, Underwood A, Chalker VJ, Spiller OB. Genomic determination of minimum multi-locus sequence typing schemas to represent the genomic phylogeny of *Mycoplasma hominis*. *BMC Genomics*. 2016;17:964.
- Adeniji JA, Faleye TOC, Adewumi OM, Olayinka OA, Donbraye E, Oluremi B, et al. Draft genome sequence of *Mycoplasma arginini* strain NGR_2017. *Genome Announc*. 2018;6:e00577–18.
- Landsberger M, Gandon S, Meaden S, Rollie C, Chevallereau A, Chabas H, et al. Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. *Cell*. 2018;174:908–e91612.
- Citti C, Dordet-Frisoni E, Nouvel LX, Kuo CH, Baranowski E. Horizontal gene transfers in mycoplasmas (*Mollicutes*). *Curr Issues Mol Biol*. 2018;29:3–22.
- Sirand-Pugnet P, Citti C, Barré A, Blanchard A. Evolution of mollicutes: down a bumpy road with twists and turns. *Res Microbiol*. 2007;158:754–66.
- Hong X, Yin J, Wang W, Zhao F, Ding X, Yu H, et al. The associations between low abundance of *Mycoplasma hominis* and female fecundability: a pregnancy-planning cohort study. *BMC Microbiol*. 2022;22:121.
- Trimomatic. <https://github.com/timflutre/trimomatic>
- FastQC. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- LongQC. <https://github.com/yfukasawa/LongQC>
- FitLong. <https://github.com/rrwick/FitLong>
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
- Spades. <https://github.com/ablab/spades>
- Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*. 2017;13:e1005595.
- Unicycler. <https://github.com/rrwick/Unicycler>
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
- Prokka. <https://github.com/tseemann/prokka>

35. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–5.
36. QUAST. <https://github.com/ablab/quast>
37. Seppy M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol*. 2019;1962:227–45.
38. Busco. <https://gitlab.com/ezlab/busco>
39. IFB core cluster. <https://www.france-bioinformatique.fr/cluster-ifb-core>
40. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9:5114.
41. Darling ACE, Mau B, Blattner FR, Perna NT, Mauve. Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14:1394–403.
42. Darling AE, Mau B, Perna NT, progressiveMauve. Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*. 2010;5:e11147.
43. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2024 June 12]. <https://www.ncbi.nlm.nih.gov/>
44. VFanalyzer. <http://www.mgc.ac.cn/cgi-bin/VFs/v5/main.cgi>. Last access 2023/12/01.
45. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res*. 2021;50:D912–7.
46. Cacciotto C, Dessi D, Cubeddu T, Cocco AR, Pisano A, Tore G, et al. MHO_0730 as a surface-exposed calcium-dependent nuclease of *Mycoplasma hominis* promoting neutrophil extracellular trap formation and escape. *J Infect Dis*. 2019;220:1999–2008.
47. Guiraud J, Le Roy C, Rideau F, Sirand-Pugnet P, Lartigue C, Bébéar C et al. Improved transformation efficiency in *Mycoplasma hominis* enables disruption of the MIB–MIP system targeting human immunoglobulins. *Microbiol Spectr*. 2023;e01873–23.
48. Lysnyansky I, Sachse K, Rosenbusch R, Levisohn S, Yogev D. The *vsp* locus of *Mycoplasma bovis*: Gene organization and structural features. *J Bacteriol*. 1999;181:5734–41.
49. Xu QY, Pan Q, Wu Q, Xin JQ. *Mycoplasma bovis* adhesins and their target proteins. *Front Immunol*. 2022;13:1016641.
50. Gaurivaud P, Baranowski E, Pau-Roblot C, Sagné E, Citti C, Tardy F. *Mycoplasma agalactiae* secretion of β -(1→6)-glucan, a rare polysaccharide in prokaryotes, is governed by high-frequency phase variation. *Appl Environ Microbiol*. 2016;82:3370–83.
51. Xie Z, Tang H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*. 2017;33:3340–7.
52. Galaxy server. <https://usegalaxy.eu/>.
53. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*. 2006;34(Database issue):D32–6.
54. ISFinder. <https://www-is.biotoul.fr/>. Last access 2023/11/30.
55. Meygret A, Peuchant O, Dordet-Frisoni E, Sirand-Pugnet P, Citti C, Bébéar C, et al. High prevalence of integrative and conjugative elements encoding transcription activator-like effector repeats in *Mycoplasma hominis*. *Front Microbiol*. 2019;10:2385.
56. Henrich B, Hammerlage S, Scharf S, Haberhausen D, Fürnkranz U, Köhrer K, et al. Characterisation of mobile genetic elements in *Mycoplasma hominis* with the description of ICEHo-II, a variant mycoplasma integrative and conjugative element. *Mob DNA*. 2020;11:30.
57. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res*. 2011;39 Web Server issue:W347–52.
58. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44:W16–21.
59. PHASTER. <http://phaster.ca/>. Last access 2023/10/22.
60. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–58.
61. Phyre2. <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>. Last access 2023/10/29.
62. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol*. 2020;21:180.
63. Panaroo. <https://github.com/gtonkinhill/panaroo>
64. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4.
65. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;gkae268.
66. iTOL. <https://itol.embl.de/>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.