Infection, Genetics and Evolution 92 (2021) 104886

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Novel prophage-like sequences in *Mycoplasma anserisalpingitidis*

Áron B. Kovács [a], Enikő Wehmann [a], Domonkos Sváb [a], Katinka Bekő [a], Dénes Grózner [a], Alexa Mitter [a], Krisztina Bali [a], Christopher J. Morrow [b], Krisztián Bányai [a], Miklós Gyuranecz [a,*]

[a] *Institute for Veterinary Medical Research, Centre for Agricultural Research, H-1143 Hungária krt. 21, Budapest, Hungary*
[b] *Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Melbourne, Victoria 3010, Australia*

ARTICLE INFO

Keywords:
In silico prophage search
Mycoplasma anserisalpingitidis
VIBRANT
Waterfowl
Whole genome

ABSTRACT

Mycoplasma anserisalpingitidis is a bacterial waterfowl pathogen. In these days of growing antibiotic resistance, it is necessary to search for alternative methods of defense against *Mycoplasma* impacts in flocks.

In order to identify prophage-like sequences, three established bioinformatics tools (PHASTER, PhiSpy, Prophage Hunter) were used in this study for the *in silico* screening of 82 *M. anserisalpingitidis* whole genomes. The VIBRANT software was used as a novel approach to further investigate the possibility of prophages in the sequences.

The commonly used softwares found prophage-like sequences in the strains, but the results were inconclusive. The VIBRANT search resulted in multiple hits, and many of them were over 10,000 base pairs (bp). These putative prophages are comparable in size to the few described mycoplasma phages. The translated coding DNA sequences of the putative prophages were checked with protein BLAST. The functions of the proteins found by the BLASTP search are common among bacteriophages. The BLASTN search of the sequences found that many of these were more similar to the *M. anatis* NCTC 10156 strain, rather than the available *M. anserisalpingitidis* strains.

The initial screening pointed at the presence of novel bacteriophages in the *M. anserisalpingitidis* and *M. anatis* strains. The VIBRANT search results were very similar to each other and none of these sequences were part of the core genome of *M. anserisalpingitidis*, with a few exceptions. The VIBRANT analysis explored presumably intact, novel prophages.

1. Introduction

Mycoplasma anserisalpingitidis is a waterfowl pathogen bacterium that was first isolated in 1983 in Hungary (Stipkovits et al., 1984) and has only recently been fully characterized as a species (Volokhov et al., 2020). It is present worldwide (Hinz et al., 1994; Gyuranecz et al., 2020) and can be isolated along with *M. anatis* from geese and occasionally ducks. The species can be part of the normal microbe community of the host but can cause significant economic losses (Stipkovits and Kempf, 1996; Stipkovits and Szathmary, 2012) during an active infection. Symptoms can include phallus and cloaca inflammation, peritonitis, salpingitis, airsacculitis, infertility of the eggs, and increased embryo lethality (Stipkovits and Kempf, 1996; Dobos-Kovacs et al., 2009; Stipkovits and Szathmary, 2012). There are no commercially available vaccines against this *Mycoplasma* species; therefore, adequate management must be maintained to prevent the manifestation of the disease in the flocks. In case of an active infection, antibiotic treatment can be used, however, this may drive the development of antibiotic resistance *via* selective pressure (Kolář et al., 2001; Bilal et al., 2018).

With the continuous growth in the number of antibiotic-resistant bacteria, it is necessary to find an alternative solution for antibiotic treatment, *e.g.* phage intervention, both in veterinary and human medicine (https://www.oie.int/en/for-the-media/amr/, https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance) as per the One Health approach. To help with this task, it is important to expand our knowledge about this bacterium. One aspect of this is the description of the phages and prophages of the *M. anserisalpingitidis*, and to create a sufficiently large database to help in the definition of further phages and prophages.

The isolation of a novel mycoplasma phage is difficult, and currently, the known viruses infecting these pathogens are not part of any established taxonomic group, with the exception of the P1 virus in the

* Corresponding author.
E-mail address: m.gyuranecz@gmail.com (M. Gyuranecz).

https://doi.org/10.1016/j.meegid.2021.104886
Received 25 December 2020; Received in revised form 16 April 2021; Accepted 26 April 2021
Available online 29 April 2021
1567-1348/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Podoviridae family (https://www.genome.jp/virushostdb/view/). Further complicating the laboratory validation of these prophages, the fact that the phages of the Mollicutes class can persist in a tissue culture, thus making the detection of the presence of phages even more difficult. Mycoplasmas and their phages show similar characteristics to animal cells and their viruses, due to the lack of cell wall, and the phages of the Mollicutes class does not always lyse the host cells but are released through budding (Putzrath and Maniloff, 1978; Putzrath et al., 1980; Dybvig et al., 2005). This phenomenon further impedes the process of mycoplasma phage isolation. This indicates that *in silico* approaches should be preferred as the first step in the identification of phages of *Mycoplasma* strains, instead of laboratory methods.

A common feature of these phages is the double-stranded (ds) DNA, with genome size varying between approximately 10 to 20 kilobase pairs (kbp). However, just as much as the members of the Mollicutes class differ from each other in size, likely, their viruses can also vary significantly in genome size. Out of the described Mollicutes infecting phages (https://www.genome.jp/kegg/), only 11 have been fully characterized, and only three of them are infecting the *Mycoplasma* genus (https://www.genome.jp/virushostdb). These low numbers are most likely due to the fact that the members of the Mollicutes class are usually fastidious, slow-growing bacteria and isolating their phages is even more difficult. The success of homology-based searches can vary greatly depending on the extent of the available database. As such, *in silico* analysis aiming at the detection of prophages has to rely more on finding sequences with the unique characteristics of phages, such as the difference in G + C content compared to the rest of the genome, the presence of attachment sites, the direction of the coding DNA sequences (CDSs) of a region, instead of homology-based searches.

There are no known avian mycoplasma phages discovered as of yet, and even described phages of the Mollicutes class are rare. The aim of this study was to screen the available *M. anserisalpingitidis* complete genomes for existing prophages, and prophage-like sequences.

There is a plethora of available *de novo* prediction tools to search for phages and phage-like sequences. Despite the abundance of tools and methods for *in silico* screening, there is no available standardized way of describing prophages in a genome. As such, a screening by multiple methods and cross-referencing of the results is recommended. In this study, we used the most established and accessible software for the analysis of 82 *M. anserisalpingitidis* and 10 *M. anatis* strains. The combined use of the PHASTER (Arndt et al., 2016) online tool, the PhiSpy (Akhter et al., 2012) software, and the Prophage Hunter (Song et al., 2019) was applied as a pilot study to determine the presence of prophage-like sequences in the genomes. The VIBRANT (Kieft et al., 2020) software searching for words in the CDS names associated with prophages was also used as a novel approach.

2. Materials and methods

2.1. Whole genome sequencing, assembly and annotation

2.1.1. Whole genome sequencing and assembly

Seventy-five *M. anserisalpingitidis* whole-genomes were provided by a previous report (Kovács et al., 2020), twenty-seven of these strains were resequenced for better coverage. Additionally nine *M. anatis* strains were sequenced for this study as seen in Supplementary Table 1. DNA was extracted from 10 ml of logarithmic-phase broth cultures of the strains using QIAamp DNA Mini kit (Qiagen Inc., Hilden, Germany) following the manufacturer's instructions. The next-generation sequencing was performed on NextSeq 500 Illumina equipment (Illumina Inc., San Diego, CA, USA), with NextSeq 500/550 High Output Kit v2.5 reagent kit (Illumina Inc.). Publicly available whole genome sequences of an additional two *M. anserisalpingitidis* and one *M. anatis* strains, ATCC: BAA-2147, and MYCAV177 and NCTC10156 (GenBank accession numbers: CP042295, CP041663, CP041664 and CP030141, respectively) were also involved in the analyses for a total of 77 *M. anserisalpingitidis* and 10 *M. anatis* genomes. The *M. anserisalpingitidis* strains were collected between 1983 and 2018, in Hungary ($n = 69$), Poland ($n = 8$). The *M. anatis* strains were collected between 1985 and 2014, in France ($n = 3$) and Hungary ($n = 7$). The draft genomes were assembled with the SPAdes program (Bankevich et al., 2012) version 3.13.0.

2.1.2. The annotation and validation of the annotation of the draft genomes

The draft genomes were annotated with RAST (Aziz et al., 2008), and the annotations were compared to the NCBI pipeline, using two available complete *M. anserisalpingitidis* genomes (ATCC BAA-2147 and MYCAV177).

2.2. The initial screening of M. anserisalpingitidis draft genomes for prophage sequences

The initial search process was divided into five stages (PHASTER, PhiSpy, Prophage Hunter, second PHASTER, secondary Prophage Hunter search) as seen in Fig. 1. This first exploratory search was done only on the *M. anserisalpingitidis* draft genomes, the *M. anatis* genomes were only used in a later step to check the possibility of a cross infection. First, the raw draft genomes (without any annotations) of the strains were checked with PHASTER (Arndt et al., 2016) online tool, which searches against an ever expanding phage database. PHASTER searches for phage-like genes and checks whether these sequences are abundant enough to be considered phage-derived. The software also assigns a score based on how many of these sequences correspond to known phage genes. If the checked region only contains phage genes, it receives the maximum score of 150, otherwise it receives a score based on the region size and the number of genes (Zhou et al., 2011). The draft genomes were annotated with RAST (Aziz et al., 2008), and the annotated sequences of the strains were then checked with PhiSpy (Akhter et al., 2012) software (version 4.1.16), which focuses on the identification of prophage characteristics in the sequences. The software checks the whole genomes of the bacteria in question based on the GC skew, the median protein length, the transcription strand orientation and the abundance of phage "words" (or twelve consecutive base pairs). The software was run with default parameters. The draft genomes were then further screened by the Prophage Hunter server. Prophage Hunter checks the transcriptional orientation, the ratio of prophage/bacterial genes, the average protein length, the composition of 20 amino acids in the genes, the Watson-Crick ratio and the transcription strand switch. The software assigns a score between 0 and 1, indicating the probability of the sequence being an active prophage, by measuring the nucleotide level genomic similarity (Song et al., 2019).

As PhiSpy uses a length filter (only uses sequences larger than 4,000 bp), and a different annotation method than the other tools, the software's output was used in a secondary check with PHASTER. The strains that gave positive results (*i.e.* the tool found prophage-like sequences) during this second check were further investigated. The draft genomes of the strains with positive hits were assembled into a single scaffold for each strain with MeDuSa (Bosi et al., 2015) online tool. These scaffolds were checked a second time with Prophage Hunter as well.

2.3. Analysis of the putative prophage sequences

The secondary PHASTER hits were analyzed with Geneious Prime software (Kearse et al., 2012) version 2019.2.3. The putative prophages were aligned with MAFFT algorithm (Katoh et al., 2002). The CDSs of these phage-like sequences were translated into protein sequences and were then subjected to a BLASTP search with NCBI blast against the non-redundant protein sequences database (Altschul et al., 1990). The Prophage Hunter results were manually added to the raw genome sequences and the PHASTER hits were mapped to these genomes with Geneious Prime 2019.2.3 (Kearse et al., 2012) for cross-examination. The putative prophages were aligned to the known mycoplasma phages, MAV1,

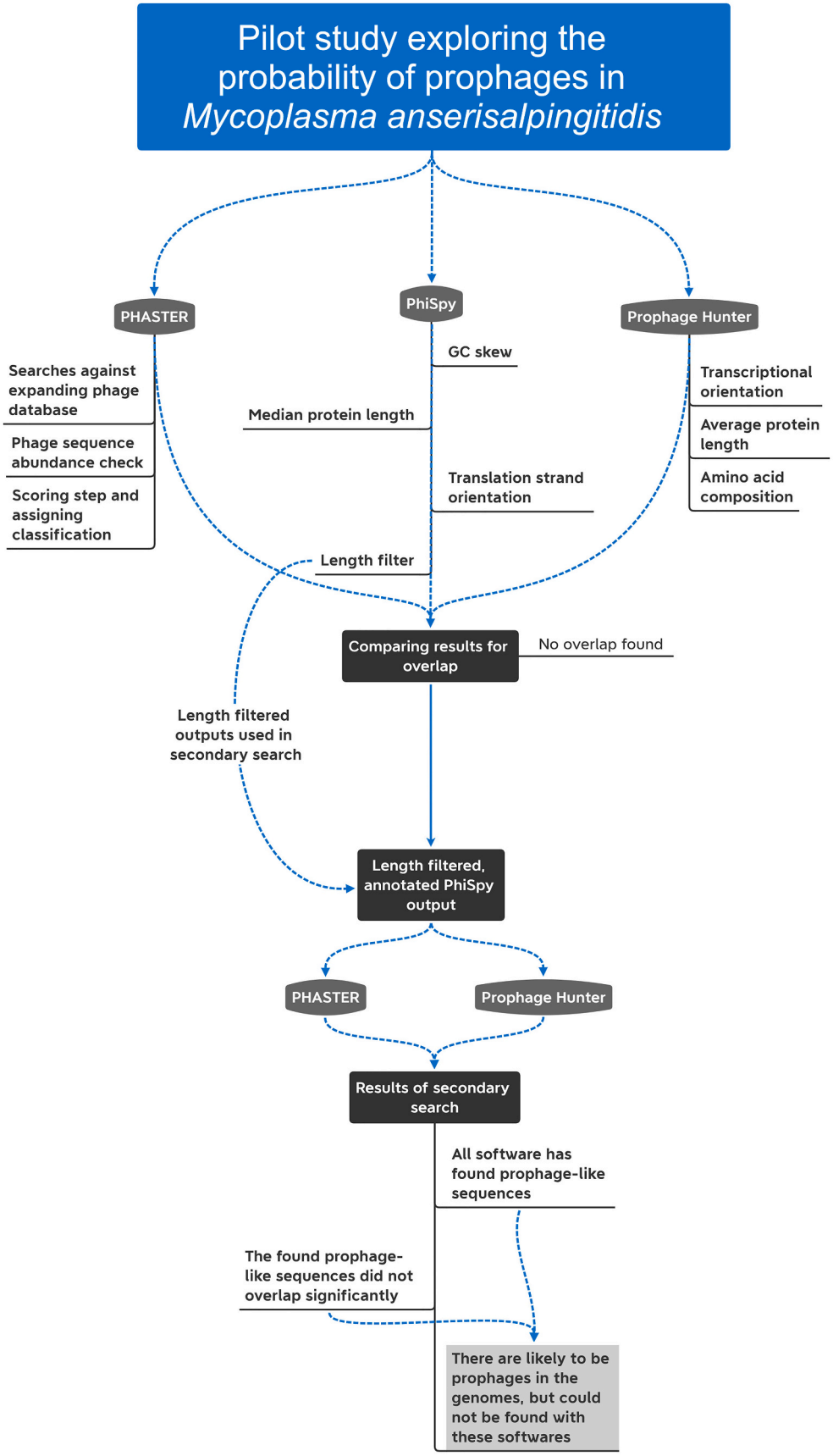

Fig. 1. The workflow of the initial screening process with three mainstream software.

phiMFV1 and P1 (Voelker and Dybvig, 1998; Tu et al., 2001; Röske et al., 2004) genomes (GenBank accession numbers: AF074945, AY583234 and NC_002515, respectively) to examine the similarity of the prophage-like sequences with these phages.

2.4. Analysis of the draft genomes with VIBRANT software

VIBRANT software (Kieft et al., 2020) (version 1.2.1) uses a novel approach for prophage searching. This tool analyzes consecutive annotations with the Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.genome.jp/kegg/), Pfam (https://pfam.xfam.org/) and Virus Orthologous Groups (VOG) (http://vogdb.org/) databases. The software not only checks for known phage genes but for proteins, which have function common in phages and prophages. PHASTER, PhiSpy and Prophage Hunter uses known phage and prophage sequences and common characteristics of prophages like GC content or gene orientation. These metrics however limit the tools in finding novel prophages in rarely studied bacteria due to the lack of reference. The putative prophages found by VIBRANT were mapped with Geneious Prime version 2019.2.3 (Kearse et al., 2012) to the respective draft genomes of the strains, and to the other strains as well, in order to check whether the putative prophages are present in them or not. The CDSs of the putative prophage sequences found with VIBRANT were blasted to the core genome multi-locus sequence typing schema of *M. anserisalpingitidis* (Kovács et al., 2020) to check for housekeeping genes (as those are unlikely to be phage sequences). Using BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011), all of the putative prophages with a length over 10,000 bp were aligned to the MeDuSa scaffold of the strain MYCAV270, which was chosen due to the large size of its prophage-like sequence. This was performed to compare the different prophage-like sequences and check for insertions or deletions. The two putative prophage sequences of MYCAV342 and MYCAV313 were clustered with the other putative prophage from their respective strain together (putative prophages MYCAV342–1 and – 2, and putative prophages MYCAV313–1 and 2). The individual CDSs of positive hits were further checked with nucleotide BLAST (Altschul et al., 1990), and the translated proteins of the putative prophage sequences with protein BLAST against the nucleotide collection and database of non-redundant protein sequences, respectively.

3. Results

3.1. Whole genome sequencing, assembly and annotation

3.1.1. Whole genome sequencing and assembly

The raw nucleotide sequence reads of the 77 *M. anserisalpingitidis* and 9 *M. anatis* strains can be found in the Sequence Read Archive (SRA) of NCBI under the BioProject number: PRJNA602215, and PRJNA682526 respectively.

3.1.2. The annotation and validation of the annotation of the draft genomes

The RAST annotation and the NCBI pipeline showed a great deal of similarity. The NCBI pipeline had 742 CDSs for MYCAV177 and 774 for the strain ATCC BAA-2147. The RAST annotation found 763 CDSs in MYCAV177 and 806 CDSs in ATCC BAA-2147. The similarity between the two annotations (based on the number of overlapping CDSs) was approximately 97.2% for MYCAV177 and 96% for ATCC BAA-2147.

3.2. The initial screening of M. anserisalpingitidis draft genomes for prophage sequences

3.2.1. PHASTER

The initial PHASTER (Arndt et al., 2016) screening resulted in a hit for all strains, however, all of these hits had a low score (between 10 and 40) making them unlikely to be actual prophages. The size of these hits did not exceed 6.9 kbp and all were ranked as incomplete. Moreover, the hits found phages from non-Mollicutes only, with the exception of MYCAV77.

3.2.2. PhiSpy

Running the PhiSpy software (Akhter et al., 2012) on the annotated draft genomes resulted in hits for all of the strains with the exception of MYCAV39, MYCAV94 and MYCAV218, however, PhiSpy found an unrealistically large number of prophage sequences in the strain MYCAV494 (36 putative prophages constituting approximately 1/3 of the genome). The prophage-like sequences found by PhiSpy were on average 11,181 bp long, between 2435 bp to 14,295 bp (comparable in size to the described mycoplasma phages).

3.2.3. Prophage Hunter

The Prophage Hunter software (Song et al., 2019) also found prophage-like sequences in all of the genomes, however, the only strains that contained mycoplasma prophages were MYCAV34, MYCAV77, MYCAV202 and MYCAV670. The probability that these sequences were phages ranged between 0.84 and 0.93 (0.89 on average).

3.2.4. Secondary PHASTER and Prophage Hunter searches

This secondary analysis resulted in only five hits (in the case of MYCAV34, MYCAV56, MYCAV77, MYCAV180 and MYCAV202), however, these were of considerably higher quality (scores varying between 80 and 90, with the exception of MYCAV34 with a score of 20) than the previous hits. All of these were found to relate to *M. arthritidis* bacteriophage MAV1. As these strains had the highest chance of containing prophages based on the results, these sequences were further analyzed. During the secondary Prophage Hunter analysis, the tool found higher quality hits, than before. Many of these related to known mycoplasma phages (phiMFV1 in MYCAV180 and MYCAV77, MAV1 in MYCAV56 and MYCAV34).

3.3. Analysis of the putative prophage sequences

During the sequence alignment of these putative prophages, it was found that four out of five were highly similar (over 80% identity in the overlapping regions) with the prophage-like sequence in MYCAV202 being very different (only 52.41% identity on average). The protein sequences from the CDSs of the putative phages were checked against the non-redundant protein sequences database with protein BLAST (Altschul et al., 1990) algorithm and resulted in hits against known mycoplasma phage (mostly MAV1) proteins, in 28 out of 68 cases of the putative prophage proteins. However, the scores were low, only 56.86% on average in coverage and 32.99% on average in the identity. The best results were 95.00% coverage and 44.88% identity (MarRP protein of phiMFV1), so less than half of the protein was similar to the BLASTP hit. Upon the cross-examination of the secondary PHASTER and Prophage Hunter results, it was found that the hits overlapped.

The alignment of the prophage-like sequences with MAV1, phiMFV1 and P1 genomes showed low identities, on average 38.89% for phiMFV1, 35.78% for P1 and 37.41% for MAV1, demonstrating that the sequences found by the software were different from these phages.

3.4. Analysis of the draft genomes with VIBRANT software

The overall abundance of positive hits with the different tools warranted the use of a novel method. The VIBRANT software (Kieft et al., 2020) was used to further investigate the strains and it found putative prophages in 37 out of the 77 strains. The length of these hits varied between 1870 and 59,948 bp, hinting at both fragmented and intact prophage-like sequences. The mapping of the sequences which were over 10,000 bp ($n = 31$) showed that these intact prophage regions were not fragmented in the genomes. Twenty-two out of the thirty-one putative prophages found in the *M. anserisalpingitidis* strains showed a great deal of similarity with *M. anatis* NCTC 10156 strain, based on separate

nucleotide BLAST of their CDs (57.00–100.00% coverage and 90.98–93.52% identity) as seen in the Supplementary Table 2. Only five putative prophages showed higher similarity to any *M. anserisalpingitidis* strains than *M. anatis* NCTC 10156 strain based on the nucleotide BLAST (between 29.00 and 100.00% coverage and 89.52–98.46% identity). The software also found prophage-like sequences in the *M. anatis* strains. Three of these sequences (in strains MYCAV786, MYCAV787 and MYCAV788) showed a high degree of similarity to the one found in MYCAV270 (approximately 90% at the overlapping regions). The cross-examination with the core genome of *M. anserisalpingitidis* revealed that only five putative prophage CDSs were found in the core genome, all of them identified in MYCAV669, proving that the rest of the sequences are part of the accessory genome. Out of all the sequences larger than 10,000 bp, twenty-one showed a great deal of similarity to each other (95.65% identical on average at the overlapping regions, approximately 15,000 bp). The BRIG output further demonstrated that these twenty-one sequences were highly similar, with the exception of insertions/deletions as seen in Fig. 2.

Ten *M. anatis* strains were also checked with VIBRANT software to see whether the putative prophages are present in other waterfowl mycoplasmas. The software found prophage-like sequences in four *M. anatis* strains and three of these putative prophages overlapped with the putative prophage found in MYCAV270 strain. Upon aligning the prophage-like sequences of the *M. anatis* strains with the one found in MYCAV270, the overlapping regions showed approximately 90.4% similarity.

The protein BLAST (Altschul et al., 1990) searches showed a high degree of similarity with proteins, which are often found in phages (DNA methylation proteins, transcription regulator proteins, transport proteins *etc.*). Numerous protein sequences of the putative prophage of MYCAV270 were hypothetical protein sequences found in the *M. anatis* NCTC 10156 strain. In some cases, the function of the protein could be theorized, based on the high degree of similarity to other known proteins. The annotated prophage-like sequence of MYCAV270 can be seen in Fig. 3.

The protein BLAST of the sequences from the *M. anatis* strains showed that the CDSs of these putative prophages mainly aligned with *M. anatis* strains. In addition, these prophage-like sequences were almost identical to each other.

The twenty-two putative prophage sequences can be found under NCBI accession numbers: MT872800 - MT872814 and MW353166 – MW353172 and MW358880.

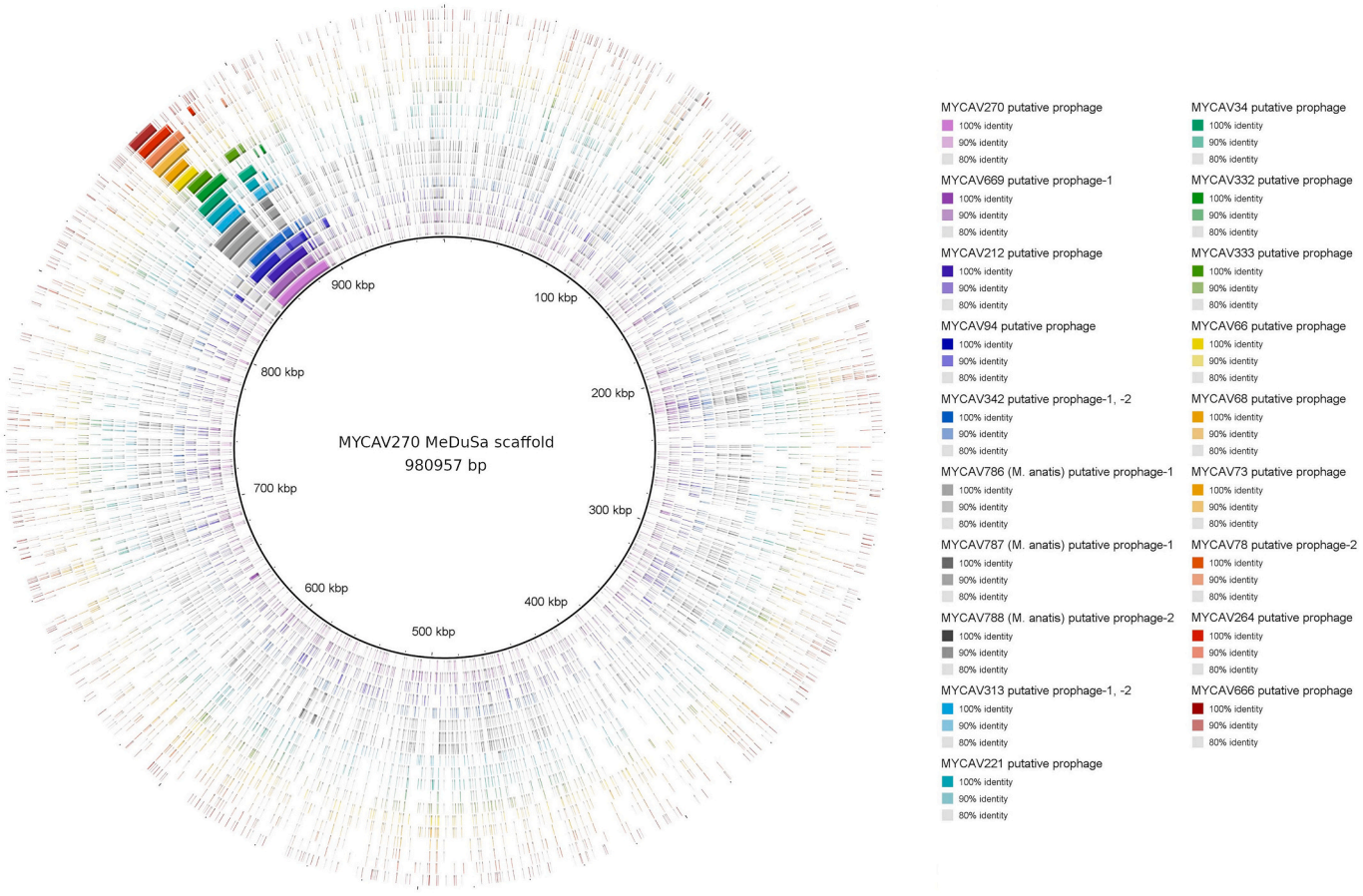

Fig. 2. The alignment of the putative prophages displaying the highest sequence identity with the scaffold MYCAV270 by BRIG.
Twenty-one prophage-like sequences (MYCAV313–1 and – 2, and MYCAV342–1 and – 2 were grouped together) were aligned to the MYCAV270 scaffold and showed a great deal of similarity (≥90%) to it and each other (the length of the putative prophage of MYCAV270 is 43,593 bp). A major insertion/deletion event can be seen in the prophage-like sequence of MYCAV221, MYCAV34 and MYCAV333 (MYCAV313–1 and – 2 showed similar profile when analyzed together) compared to the putative prophage of MYCAV270. This insertion/deletion can also be seen in the *M. anatis* strains in gray (MYCAV786–788). Eight putative prophage only align with half of the MYCAV270 putative prophage. The nucleotide position in the MYCAV270 MeDuSa Scaffold is not certain, as this scaffold was done with the help of reference genomes.

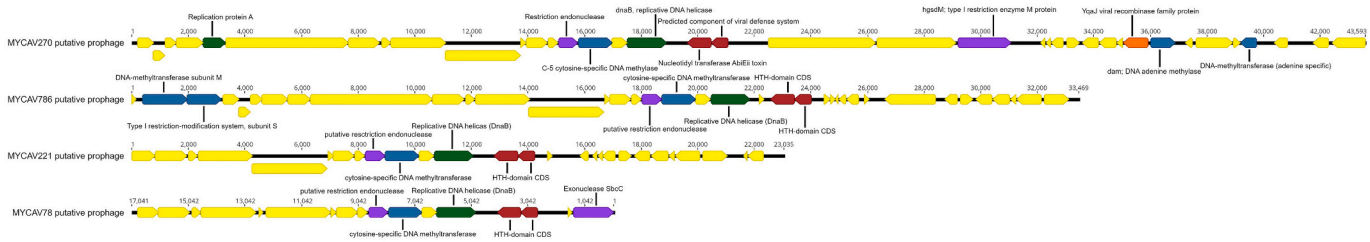

Fig. 3. The annotated prophage-like sequence of MYCAV270 and three other representative putative prophages (from MYCAV221, MYCAV78 and *M. anatis* strain MYCAV786).
The yellow colored arrows denote hypothetical proteins or proteins with unknown function in bacteria. Green color represents CDSs part of the replication process, purple are restriction enzymes/endonucleases, blue are methylases or methyltransferases, red are part of the viral defense system, while orange is a viral recombinase. The sequences are not aligned, instead are shown from start to finish (hence insertions/deletions can not be ascertained from the image).(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Mycoplasma anserisalpingitidis is a common pathogen of geese and can be found in ducks as well. It can cause significant economic damage in the waterfowl industry. There are no commercially available vaccines against this species as of yet, and as all mycoplasmas, it has an innate lack of susceptibility against multiple antibiotics (like β-lactams due to the lack of cell wall) (Chernova et al., 2016). In the case of *M. anserisalpingitidis*, elevated minimum inhibitory concentration values have been reported for multiple antibiotics, like spectinomycin, macrolides or lincomycin (Grózner et al., 2016). This further restricts the number of antibiotics applicable to these bacteria. With the increase of acquired antibiotic-resistance, it is imperative to map the genetic variance of the antibiotic resistant bacteria. An important part of understanding a pathogen is the proper description of the genetic material, including any potential prophages. Phages and prophages can be vectors of horizontal gene transfer (HGT) and can contribute significantly to the genetic variability of the strains. In a HGT event strains or species can exchange genetic materials, such as genes coding for antibiotic resistance mechanisms.

Our knowledge is very limited regarding the bacteriophages of the Mollicutes class and mycoplasmas in general. Only three mycoplasma infecting phages have been fully characterized, the MAV1, the phiMFV1, and the P1 phage (Röske et al., 2004; Tu et al., 2001; Voelker and Dybvig, 1998), and none have been described in avian mycoplasmas. In a recent study a GC rich prophage like region have been identified in the genome of *M. bovirhinis* strain HAZ141_2 (Lysnyansky and Borovok, 2021), however this region showed very little similarity to the MYCAV270 putative prophage like sequence (only 37.33% similarity based on MAFFT alignment). In this study, we used the freely available bioinformatics tools from well-established methods to novel approaches, to screen the full genomes of *M. anserisalpingitidis* strains for potential phages. Due to the lack of substantial mycoplasma phage genomes, the *in silico* analysis had to rely mostly on the detection of common characteristics of prophages, such as the presence of attachment sites, the G + C content, or the direction of the CDSs. The fact that all of the softwares used in this study found results suggests the presence of unknown prophage- or phage-like regions.

Due to the lack of established methods for prophage search, multiple tools were used in this study, but the results only marginally supported each other. One reason behind this can be the fact that each software uses a slightly different approach. The PHASTER (Arndt et al., 2016) online available tool searches for phages *via* Glimmer (Salzberg et al., 1998) annotation (in raw nucleotide sequences) and checks if the possible phage-like sequences are sufficiently close together to be considered a prophage. It has a constantly (biweekly) updated phage database used in the comparison of the annotated proteins. On the other hand, the PhiSpy tool (Akhter et al., 2012) uses annotated sequences (in this study by the RAST annotation software (Aziz et al., 2008)). The program checks for prophage-like characteristics in the genome, like the length of the coded proteins, the direction of the CDSs, the AT/GC skew of the region compared to the rest of the genome, or phage insertion points. The third established method used in this study was the Prophage Hunter (Song et al., 2019). The software combines the comparison-based methods, the attachment (att) site search, and a machine-learning algorithm to predict the presence of a prophage. The final software used was the VIBRANT (Kieft et al., 2020) tool, a novel phage searching method based on combining machine learning and multiple annotation filtering steps. Concisely, the software searches for words in the CDS names associated with prophages and checks if the positive hits are in sufficient abundance and proximity to form a potential prophage.

The most prevalent tools (PHASTER, PhiSpy and Prophage Hunter) found prophage-like sequences in all of the strains; however, cross-examination revealed that these positive hits often differed greatly from each other. A reason behind non-overlapping positive hits can be the fact that both *M. anserisalpingitidis* and mycoplasma phages, in general, are rarely studied. The PHASTER tool and the Prophage Hunter software gave similar outcome, but upon checking the result with nucleotide and protein BLAST, these hits were not found to be related to described mycoplasma phages, nor any known phage-like proteins. The results however pointed at the presence of novel prophages in the strains.

The outcome of the VIBRANT software shows great promise, insofar as the finding of potential novel prophages are concerned. The hits varied greatly in length, but there were multiple prophage-like sequences over 10,000 bp, with some reaching 20,000 or even over 40,000 bp. The comparison between these lengths and the size of the described mycoplasma phages shows a great deal of similarity (Voelker and Dybvig, 1998; Tu et al., 2001; Röske et al., 2004). The nucleotide BLAST result showed that most of these phage-like sequences were present in the *M. anatis* NCTC 10156 strain, but not in the *M. anserisalpingitidis* ATCC:BAA-2147 strain, hinting at a possible phage infection or horizontal gene transfer in the past. The two species are closely related and show a high degree of similarity with each other (over 99% in 16S rRNA genes) (Volokhov et al., 2020). The fact that many of these sequences had major insertions/deletions, but were otherwise very similar, serves as another proof of the prophagic origin of these sequences. The GC content and the codon usage of the putative prophages and the host strains have been checked with Geneious Prime 2019.2.3 and the Sequence Manipulation Suite (Stothard, 2000) and no significant difference have been found in the GC content and codon usage of the putative prophage sequences and the host sequences.

The VIBRANT tool proved to be valuable in the screening of the strains especially since it is not restricted to comparison with known phages and prophages and does not rely on genetic markers like GC content or gene orientation. The fact that the genes of the potential prophages were not part of the *M. anserisalpingitidis* core genome, and as such are not housekeeping genes, further points toward the possibility of

these sequences being prophages. The coded proteins are common among bacteriophages, like methyltransferases, phage resistance proteins, integrases, proteins in intracellular transport, transcription regulators and so on. The lack of structural proteins complicate the description of these putative prophages, however, it is important to note that many hypothetical proteins with no described function have been found. Mobilization experiments and characterization of their infective properties with suitable host strains are needed to assess the actual role of these prophages in the genomics, physiology, and ecology of *M. anserisalpingitidis* strains.

M. anatis strains also had putative prophages highly similar to the prophage-like sequences found in the *M. anserisalpingitidis* strains, however, these sequences were not present in all the strains of the two species, further suggesting that these are mobile elements in waterfowl mycopalsmas. Volokhov et al. (2020) hypothesized that the close relation between *M. anserisalpingitidis* and *M. anatis* are due to their evolution from a common ancestor. This hypothesis can also serve as an explanation to the similar prophage-like sequences, as these putative prophages could have originated from this common ancestor bacterial species. Another possible reason behind the similarities is that these sequences crossed over in the past from one *Mycoplasma* species to the other.

5. Conclusions

While laboratory evidence is missing for the complete characterization of these putative prophages, the result shows a great deal of promise. Our findings can be considered as circumstantial evidence that these sequences are actually prophages. The prophage-like sequences found in this study can help with the establishment of a mycoplasma phage database. These putative prophages can serve as a base for other studies aiming at detecting mycoplasma phages and can help with the discovery of the genetic variance and can be the cause of genetic changes.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.104886.

Ethics approval and consent to participate

According to the written declaration (reference number: IVMR/2019/0023) of the Ethics Committee of the Institute for Veterinary Medical Research, Centre for Agricultural Research ethical approval was not required for the study as the samples were taken during routine diagnostic examinations with the written consent of the owner. The final study and the manuscript was submitted and approved in a written declaration by the Ethics Committee of the Institute for Veterinary Medical Research, Centre for Agricultural Research.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article. The sequence reads of the *M. anatis* strains have been uploaded to the Sequence Read Archive (SRA) under PRJNA682526. The twenty-one prophage-like sequences have been uploaded to GenBank and can be found under the accession numbers: MT872800 - MT872814, MW353166 - MW353172 and MW358880, the prophage like sequences have also been uploaded to the Figshare website, along with the draft genomes. The putative prophage sequences can be found at: https://figshare.com/articles/dataset/Waterfowl_mycoplasma_putative_prophages/14034554. The draft genomes of the *Mycoplasma anserisalpingitidis* strains can be found at: https://figshare.com/articles/dataset/Mycoplasma_anserisalpingitidis_draft_genomes/13855091 and the *M. anatis* draft genomes can be found at: https://figshare.com/articles/dataset/Mycoplasma_anatis_draft_genomes/14034548.

Funding

This work was supported by the Lendület program (LP2012-22) of the Hungarian Academy of Sciences, the K_16 (119594), and KKP19 (129751) grants of the National Research, Development and Innovation Office, Hungary. MG was supported by the Bolyai János Fellowship of the Hungarian Academy of Sciences and by the Bolyai+ Fellowship (ÚNKP-19-4-ÁTE-1) of the New National Excellence Program of the Ministry of Innovation and Technology. KÁB was supported by the Cooperative Doctoral Programme for Doctoral Scholarships (KDP-2020) of the National Research, Development and Innovation Office, Hungary. DS was supported by the Bolyai János Fellowship of the Hungarian Academy of Sciences. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

Declaration of Competing Interest

The authors declare that they have no competing interests.

References

Akhter, S., Aziz, R.K., Edwards, R.A., 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. Nucleic Acids Res. 40, 1–13. https://doi.org/10.1093/nar/gks406.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12. https://doi.org/10.1186/1471-2164-12-402.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., Wishart, D.S., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44, W16–W21. https://doi.org/10.1093/nar/gkw387.

Aziz, R.K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9, 1–15. https://doi.org/10.1186/1471-2164-9-75.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Bilal, A., Wang, W., Arshad, M.I., Khurshid, M., Rasool, M.H., Nisar, M.A., Aslam, M.A., Qamar, M.U., Salamat, M.K.F., Baloch, Z., 2018. Antibiotic resistance: a rundown of a global crisis. Infect. Drug Resist. 1645–1658. https://doi.org/10.2147/IDR.S173867.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.F., Lió, P., Crescenzi, P., Fani, R., Fondi, M., 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31, 2443–2451. https://doi.org/10.1093/bioinformatics/btv171.

Chernova, O.A., Medvedeva, E.S., Mouzykantov, A.A., Baranova, N.B., Chernov, V.M., 2016. Mycoplasmas and their antibiotic resistance: the problems and prospects in controlling infections. Acta Nat. 8, 24–34 doi:10.32607/20758251-2016-8-2-24-34.

Dobos-Kovacs, M., Varga, Z., Czifra, G., Stipkovits, L., 2009. Salpingitis in geese associated with mycoplasma sp. strain 1220. Avian Pathol. 38, 239–243. https://doi.org/10.1080/03079450902912127.

Dybvig, K., Tu, A., Clapper, B., 2005. Mycoplasma phages. In: Waldor, M.K., Friedman, I. D., Adhya, S.L. (Eds.), Phages: Their Role in Bacterial Pathogenesis and Biotechnology, pp. 223–237. https://doi.org/10.1128/9781555816506.ch11.

Grózner, D., Kreizinger, Z., Sulyok, K.M., Rónai, Z., Hrivnák, V., Turcsányi, I., Jánosi, S., Gyuranecz, M., 2016. Antibiotic susceptibility profiles of Mycoplasma sp. 1220 strains isolated from geese in Hungary. BMC Vet. Res. 12, 1–9. https://doi.org/10.1186/s12917-016-0799-0.

Gyuranecz, M., Mitter, A., Kovács, Á.B., Grózner, D., Kreizinger, Z., Bali, K., Bányai, K., Morrow, C.J., 2020. Isolation of Mycoplasma anserisalpingitidis from swan goose (*Anser cygnoides*) in China. BMC Vet. Res. 16, 178. https://doi.org/10.1186/s12917-020-02393-5.

Hinz, K.-H., Pfützner, H., Behr, K.-P., 1994. Isolation of Mycoplasmas from clinically healthy adult breeding geese in Germany. J. Vet. Med. 41, 145–147. https://doi.org/10.1111/j.1439-0450.1994.tb00217.x.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kieft, K., Zhou, Z., Anantharaman, K., 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome 8, 90. https://doi.org/10.1186/s40168-020-00867-0.

Kolář, M., Urbánek, K., Látal, T., 2001. Antibiotic selective pressure and development of bacterial resistance. Int. J. Antimicrob. Agents 17, 357–363. https://doi.org/10.1016/S0924-8579(01)00317-X.

Kovács, Á.B., Kreizinger, Z., Forró, B., Grózner, D., Mitter, A., Marton, S., Bali, K., Sawicka, A., Tomczyk, G., Bányai, K., Gyuranecz, M., 2020. The core genome multilocus sequence typing of Mycoplasma anserisalpingitidis. BMC Genomics 21, 403. https://doi.org/10.1186/s12864-020-06817-2.

Lysnyansky, I., Borovok, I., 2021. A GC-rich prophage-like genomic region of mycoplasma bovirhinis HAZ141_2 carries a gene cluster encoding resistance to kanamycin and neomycin. Antimicrob. Agents Chemother. https://doi.org/10.1128/AAC.01010-20.

Putzrath, R.M., Maniloff, J., 1978. Properties of a persistent viral infection: possible lysogeny by an enveloped nonlytic mycoplasmavirus. J. Virol. 28, 254–261. https://doi.org/10.1128/jvi.28.1.254-261.1978.

Putzrath, R.M., Cadden, S.P., Maniloff, J., 1980. Effect of cell membrane composition on the growth and composition of a nonlytic enveloped Mycoplasmavirus. Virology 106, 162–167. https://doi.org/10.1016/0042-6822(80)90235-4.

Röske, K., Calcutt, M.J., Wise, K.S., 2004. The mycoplasma fermentans prophage φMFV1: genome organization, mobility and variable expression of an encoded surface protein. Mol. Microbiol. 52, 1703–1720. https://doi.org/10.1111/j.1365-2958.2004.04087.x.

Salzberg, S.L., Deicher, A.L., Kasif, S., White, O., 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 26, 544–548. https://doi.org/10.1093/nar/26.2.544.

Song, W., Sun, H.X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., Wang, D., Wang, Y., Hu, M., Liu, W., Yang, H., Shen, Y., Li, J., You, L., Xiao, M., 2019. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res. 47, W74–W80. https://doi.org/10.1093/nar/gkz380.

Stipkovits, L., Kempf, I., 1996. Mycoplasmoses in poultry. Rev. Sci. Tech. 15, 1495–1525.

Stipkovits, L., Szathmary, S., 2012. Mycoplasma infection of ducks and geese. Poult. Sci. 91, 2812–2819. https://doi.org/10.3382/ps.2012-02310.

Stipkovits, L., Varga, Z., Dobos-Kovacs, M., Santha, M., 1984. Biochemical and serological examination of some Mycoplasma strains of goose origin. Acta Vet. Hung. 32, 117–125.

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

Tu, A.H.T., Voelker, L.R.L., Shen, X., Dybvig, K., 2001. Complete nucleotide sequence of the mycoplasma virus P1 genome. Plasmid 45, 122–126. https://doi.org/10.1006/plas.2000.1501.

Voelker, L.R.L., Dybvig, K., 1998. Characterization of the lysogenic bacteriophage MAV1 from Mycoplasma arthritidis. J. Bacteriol. 180, 5928–5931. https://doi.org/10.1128/jb.180.22.5928-5931.1998.

Volokhov, Dmitriy V., Grózner, D., Gyuranecz, M., Ferguson-Noel, N., Gao, Y., Bradbury, J.M., Whittaker, P., Chizhikov, V.E., Szathmary, S., Stipkovits, L., 2020. Mycoplasma anserisalpingitidis sp. nov., isolated from European domestic geese (*Anser anser* domesticus) with reproductive pathology. Int. J. Syst. Evol. Microbiol., ijsem004052 https://doi.org/10.1099/ijsem.0.004052.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39, 347–352. https://doi.org/10.1093/nar/gkr485.