



# OPEN Genome-wide association study of *Mycoplasma anserisalpingtonis* strains for antibiotic susceptibility

Áron B. Kovács<sup>1,2,3</sup>, Enikő Wehmann<sup>1,2</sup>, Katinka Bekő<sup>1</sup>, Dénes Gróznér<sup>1</sup>, Krisztina Bali<sup>1</sup>, Zsuzsa Kreizinger<sup>1</sup>, Anna Sawicka<sup>4</sup>, Krisztián Bányai<sup>1,2,5</sup> & Miklós Gyuranecz<sup>1,2,5,6</sup>✉

*Mycoplasma anserisalpingtonis* is a facultative pathogenic bacterium affecting waterfowl, predominantly geese and sporadically ducks. Understanding the molecular basis of antimicrobial resistance mechanisms is crucial in the preservation of antibiotic efficiency. This study aimed to elucidate the genetic background of antibiotic susceptibility profiles of 110 *M. anserisalpingtonis* strains against nine antimicrobial agents. Significant associations between k-mers and five (tylvalosin, tilmicosin, enrofloxacin, lincomycin, spectinomycin) of the nine antimicrobial agents were identified by pyseer. Significant associations were found in multiple coding sequences that encode various members of efflux pumps, epigenetic regulation and topoisomerases among many other groups of functions. Certain k-mers associated with genes found putative prophage-like sequences suggest potential horizontal gene transfer events that could facilitate the acquisition of novel resistance mechanisms. Based on our findings, the genetic background of antimicrobial resistance of *M. anserisalpingtonis* is composed of multiple factors. Our results not only correlated with the majority of known antibiotic resistance mechanisms (e.g. drug target modification, efflux pumps, methyltransferases) but also showed potentially novel genes that could play a significant role in antimicrobial resistance. The results may serve to expedite the diagnosis of *M. anserisalpingtonis* antibiotic susceptibility profiles and support the fight against the spreading of resistance.

**Keywords** Waterfowl, Whole genome analysis, Antibiotic susceptibility, Macrolide, Lincosamide, Tetracycline, Aminocyclitol, Fluoroquinolone, Pleuromutilin

*Mycoplasma anserisalpingtonis* is a waterfowl facultative pathogenic bacterium, primarily found in geese and occasionally infecting ducks<sup>1</sup>. Mycoplasmas lack cell wall and are among the smallest known prokaryotes in terms of genome size. *Mycoplasma anserisalpingtonis* was first isolated in 1983 in Hungary<sup>2</sup> and was fully characterized in 2020 by Volokhov et al.<sup>1</sup>. This pathogen has been frequently detected in Hungarian geese farms<sup>3</sup>, and it is probably present around the whole world<sup>4,5</sup>, as it has been found in geese in Poland, Ukraine and the Russian Federation<sup>6</sup>, China<sup>4</sup> and Vietnam. *Mycoplasma anserisalpingtonis* can be part of the host's normal microbiome; however, clinical manifestation usually occurs due to different stress factors, such as cold weather, inappropriate housing conditions, or sexual activity. Clinical signs can include infertility of eggs, increased embryo lethality, phallus and cloaca inflammation, peritonitis, salpingitis, and airsacculitis, leading to high economic losses<sup>7-9</sup>.

There are multiple ways of combating *Mycoplasma* infections, like prevention by keeping the flock pathogen-free or vaccinating against the bacteria; however, there are currently no commercially available vaccines against *M. anserisalpingtonis*. In case of a disease occurrence, antimicrobial treatment is the primary option to be used against *M. anserisalpingtonis*.

There is a well-documented increase in resistance toward antimicrobial agents in both human and veterinary medicine. This underscores the necessity of rigorous antimicrobial susceptibility testing to expedite targeted antimicrobial therapy. One commonly used method for antimicrobial susceptibility testing in mycoplasmas is the broth microdilution method<sup>10</sup>; however, it might require multiple weeks. Such a prolonged period without treatment can lead to significant economic losses and demands the development of rapid diagnostic tests.

<sup>1</sup>HUN-REN Veterinary Medical Research Institute, Hungária Krt. 21, Budapest 1143, Hungary. <sup>2</sup>National Laboratory of Infectious Animal Diseases, Antimicrobial Resistance, Veterinary Public Health and Food Chain Safety, Hungária Krt. 21, Budapest 1143, Hungary. <sup>3</sup>Centre for Metagenomics, University of Debrecen, Debrecen, Hungary. <sup>4</sup>Department of Poultry Diseases, National Veterinary Research Institute, Aleja Partyzantów 57, 24-100 Pulawy, Poland. <sup>5</sup>University of Veterinary Medicine, H-1078 István U. 2, Budapest, Hungary. <sup>6</sup>Molliscience Kft., 2051 Március 15. Utca 1., Biatorbágy, Hungary. ✉email: m.gyuranecz@gmail.com

Multiple environmental factors can drive the dissemination of antimicrobial resistance mechanisms, including selective pressure due to antibiotic therapy or horizontal gene transfer between microorganisms. Variable mechanisms of antimicrobial resistance exist also, depending on the antimicrobial agents, like DNA mutations, increased efflux pump activity or methylation<sup>11</sup>. In-depth genomic analyses may enable researchers to find the link between genotype and phenotype regarding antibiotic resistance and support the development of new diagnostic methods for antibiotic susceptibility testing by designing assays targeting these genetic markers. Several studies analyzed the antibiotic susceptibility in *M. anserisalpingtonis* examining single nucleotide polymorphisms (SNPs), biofilm formation and efflux pump mechanisms<sup>12–14</sup>.

One method of contrasting phenotype and genomic markers is to conduct genome-wide association study (GWAS). Genome-wide association studies have been extensively used to dissect the genomic background of many human diseases, like myocardial infarction<sup>15</sup> and age-related macular degeneration<sup>16</sup>; however, since these early GWA studies, this method has been applied to microbial phenotypes, such as virulence factors or antimicrobial resistance<sup>17–20</sup>. An advantage of this approach is that it cannot only confirm previously described mutations but can also uncover novel mutations, potentially outside the coding DNA sequences (CDSs) of the bacterial genome, which can help in the development of new diagnostic methods. However, these studies require a large number of bacterial genomes, complete and accurate genome sequences for these, and comprehensive phenotypic data.

As GWA studies use various statistical methods (e.g. Fishers exact test, binomial test, fixed effect generalized linear regression); therefore, the number of samples necessary for the experiment to be successful depends on how strongly a certain phenotype in question correlates with the genotype. This number can vary from approximately a hundred to thousands of samples, with more samples allowing the detection of genetic variances with smaller impacts. With advances in whole genome sequencing, it has become ever easier and cheaper to attain reliable whole-genome sequences (WGS) of different bacteria. As mentioned prior it is important for the sequencing to be precise and the coverage of the sequences to be deep and high standard regarding the Phred score (indicates the measure of the base quality) and the depth (describes how often a given nucleotide has been read in the experiment) of the sequencing should be set up. The third important factor in GWAS is phenotypic data, which can range from binary characteristics (such as manifestation of an illness in animals and humans or the ability to metabolize certain molecules) to continuous (like the physical parameters of the organisms or antimicrobial resistance of a microbe). Certain characteristics are relatively easy to note (e.g. host species, physical parameters), while others have to be acquired through various methods (e.g. antimicrobial resistance profile). There are multiple ways of associating genome sequences with phenotypes, the most prominent ones are checking associations between SNPs; k-mers (“k” length short sequences), or unities<sup>21</sup>. All of the aforementioned markers have been well demonstrated in different studies<sup>22–27</sup>.

Despite the widespread usage of GWAS in human medicine and their obvious benefits, microbial GWA studies still heavily lag behind examinations conducted on human illnesses. The first GWA studies have been performed on mycoplasmas only recently (2020), all analyzing *M. bovis*<sup>28–30</sup>. In this study, we aimed to dissect the genetic background of antibiotic susceptibility profiles of 110 *M. anserisalpingtonis* strains against nine antimicrobial agents.

## Results

### Antimicrobial susceptibility testing

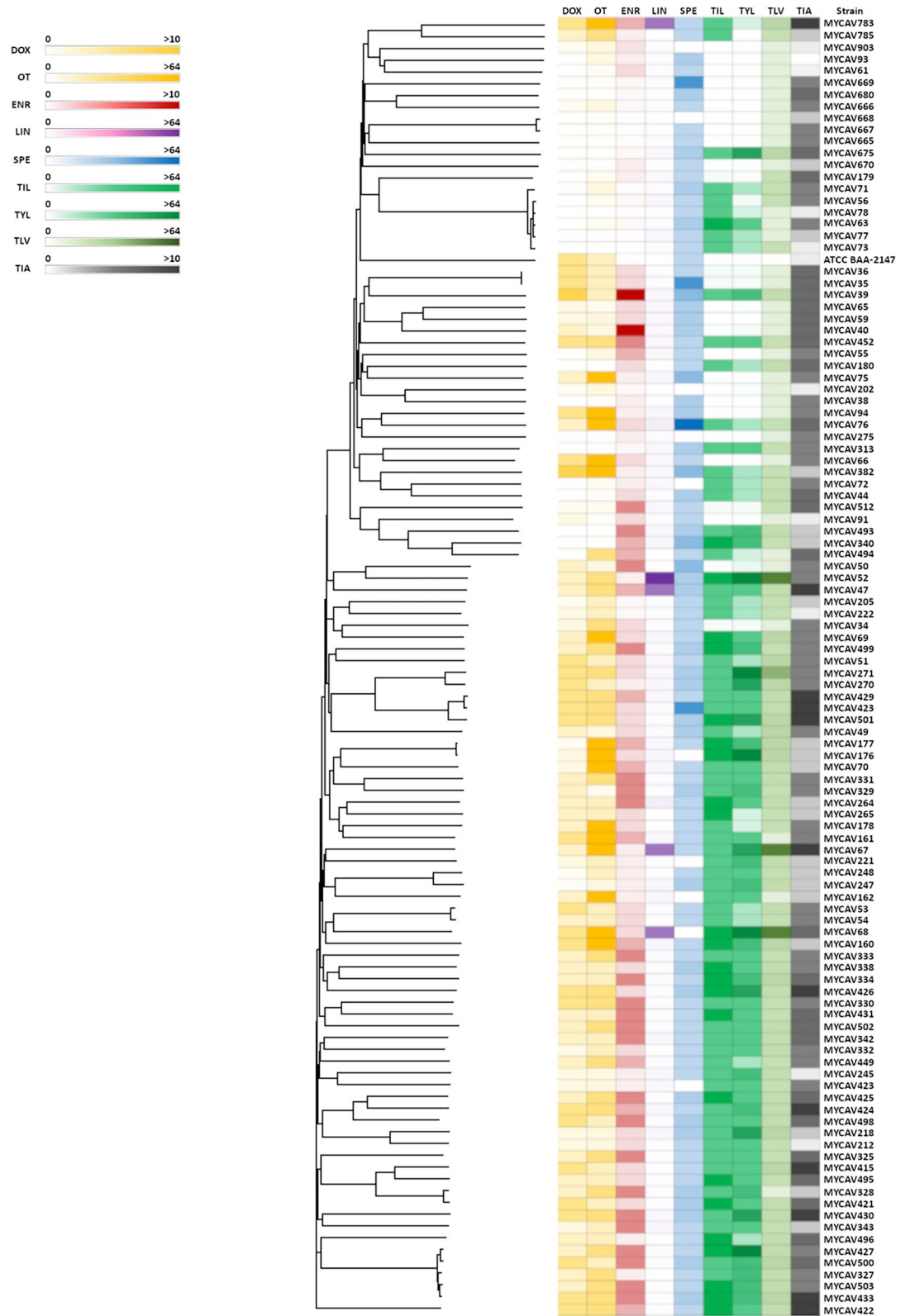
The quality control strain (ATCC BAA-2147; CP042295) showed consistent results throughout the study. Strains with elevated minimum inhibitory concentration (MIC) values were found in the cases of all tested antibiotics. Among the three macrolides investigated in this study a wide range of MIC values could be observed, with notably high MIC<sub>50</sub> and MIC<sub>90</sub> values recorded for tilmicosin ( $\geq 64$  µg/ml for both MIC<sub>50</sub> and MIC<sub>90</sub>). Tylosin inhibited at 16 µg/ml the 50% of the strains and at  $\geq 64$  µg/ml the 90% of the strains, while tylvalosin showed the lowest MIC<sub>50</sub> value (0.5 µg/ml) and MIC<sub>90</sub> value (2 µg/ml) against the strains. The MIC values for lincomycin clustered around the MIC<sub>50</sub> value (2 µg/ml), although one isolate yielded high MIC values ( $> 64$  µg/ml). Tetracyclines displayed broad ranges of the MIC values ( $\leq 0.25$  to  $> 64$  µg/ml for oxytetracycline and 0.078 to  $> 10$  µg/ml for doxycycline), with high MIC<sub>50</sub> (16 and 5 µg/ml, respectively) and MIC<sub>90</sub> values (64 and 10 µg/ml, respectively). Spectinomycin exhibited a MIC<sub>50</sub> of 8 µg/ml, with the majority of strains displaying MIC<sub>50</sub> or higher values. The MIC values for enrofloxacin showed a wide range (1.25 to  $> 10$  µg/ml), with MIC<sub>50</sub> and a MIC<sub>90</sub> values above 10 µg/ml. Tiamulin also showed a wide range of MIC values between 0.078 and 2.5 µg/ml. The MIC<sub>50</sub> value of tiamulin was 0.625 µg/ml, while the MIC<sub>90</sub> was 2.5 µg/ml. Individual MIC values against each strain are given in Supplementary Table 1, while MIC value graphs are given in Supplementary Fig. 1.

### Whole genome sequencing and assembly

The short reads from previous studies can be found under PRJNA602215, PRJNA650261 and PRJNA602206. While the sequences from this study can be found under PRJNA856806. Based on the whole genome sequencing from the previous studies<sup>3,31–33</sup>, the average number of short reads was 4 753 178 (Minimum: 739 708, Maximum: 30 383 838) with read lengths between 35 and 151 base pairs. The average Phred score was 35.4 (over 99.9% chance of correct base call) (Minimum: 31.9, Maximum: 35.4). The coverage of the genomes was on average 493.9× (Minimum: 51.7×; Maximum 2268.3×). The average length of the genomes after MAUVE alignment was 943 143 base pairs (bps) (Minimum: 862 389 bps, Maximum: 1 097 017 bps).

### Genotype: phenotype association test and promoter mapping

The cgMLST phylogenetic tree and the antimicrobial profiles are presented in Fig. 1. The pyseer tool found various numbers of significant ( $p$ -value  $\leq 5 \times 10^{-8}$ ) k-mers for 5 of the 9 antimicrobial agents. No k-mer showed significant association with tiamulin resistance neither with tetracycline (oxytetracycline or doxycycline)



**Fig. 1.** The cgMLST phylogenetic tree of all of the studied *M. anserisalpiginidis* strains along with their antibiotic resistance profile for all of the studied antimicrobial agents. The individual scales for the antimicrobial agents are on the left side of the image (the MIC values are given in  $\mu\text{g/ml}$ ). Abbreviations: DOX—doxycycline, OT—oxytetracycline, ENR—enrofloxacin, LIN—lincomycin, SPE—spectinomycin, TIL—tilmicosin, TYL—tylosin, TLV—tylvalosin, TIA—tiamulin.

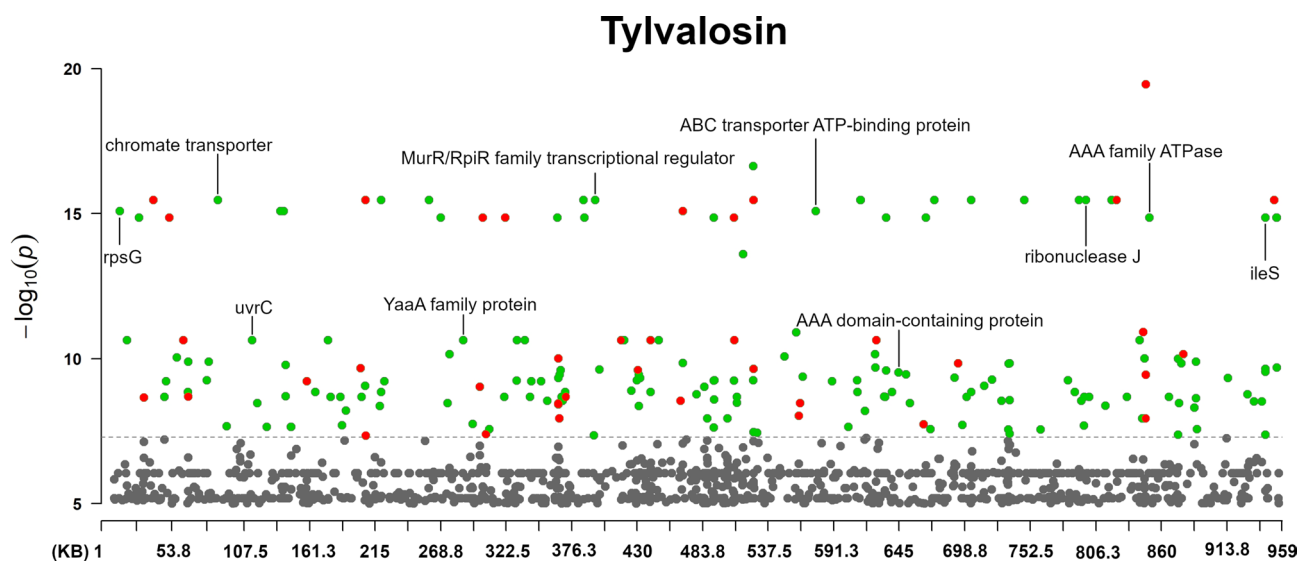
resistance. For the macrolides no significant k-mer was found for tylosin, but there were significant hits in tylvalosin and tilmicosin (92 and 727 k-mers, respectively). All in all, 19 significant k-mers were found for enrofloxacin, 1 173 significant k-mers for lincomycin and 109 significant k-mers for spectinomycin with pyseer. Promotech found 3 437 promoters in the genome of BAA-2147 type strain.

The Manhattan plots of the antibiotic agents are shown in Figs. 2, 3, 4, 5, and 6 (enrofloxacin, lincomycin, spectinomycin, tilmicosin and tylvalosin, respectively).

The most common types of CDS with a significant hit were coding for different transferases ( $n=17$ ), especially methyltransferases ( $n=7$ ). These were followed by CDSs encoding proteins with efflux pump activity or members of the efflux pump ( $n=16$ ). The third most numerous CDSs with positive hits were those coding for ligases and proteins involved in DNA repair ( $n=10$  in both cases). It is important to note that there were multiple CDSs encoding hypothetical proteins or proteins with unknown or poorly characterized functions. In the case of lincomycin we found significant k-mers in the abortive phage infection protein CDS and the YqJ viral recombinase family protein CDS, both of which was also found in a previous putative prophage study on *M. anserisalpinitidis*<sup>32</sup>. The function of the CDSs with significant k-mer associations are summarized in Table 1 and detailed in Supplementary dataset 1 (available at <https://doi.org/10.6084/m9.figshare.25378600>).

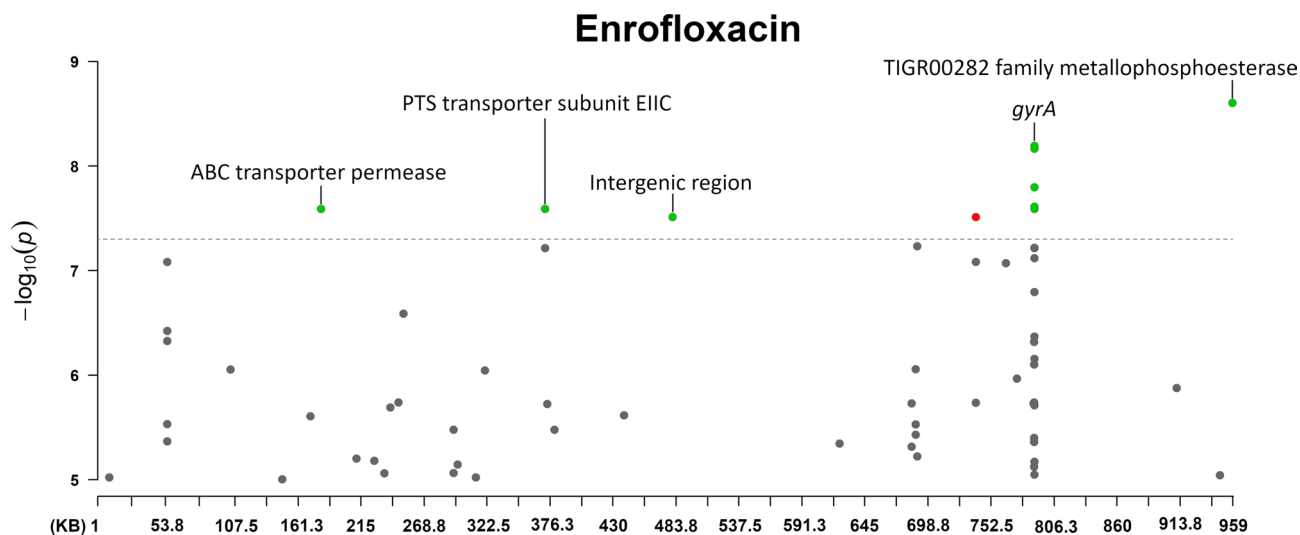
## Discussion

Multiple things can influence the outcome of a genome-wide association study, like the population structure, as the closer the relationship between certain samples or strains the higher the similarities in their genome. Another issue can be the distribution of phenotypic data, if it is skewed too much associating k-mers or SNPs with the phenotype can be problematic. The size of the genome can also influence the necessary sample size of a GWAS, as longer genomes have more genes and thus lead to more statistical comparison and would necessitate a higher sample size. Also, GWAS is only capable of dissecting the genetic background of a trait and unable to account for non-genetic elements of a phenotype, like methylation pattern or small non coding RNAs<sup>34</sup>. These factors can hinder the effectiveness of GWAS, as they are contributing to the phenotype, but not being detectable with DNA sequencing. These can also be the cause behind our study not finding significant k-mers in tiamulin, tetracyclines or in tylosin. As certain resistance associated-genes could be transmitted between pathogens, and bacteria which can infect both animals and humans may develop resistance in the animals and could be transmitted to humans via the food chain, it is extremely important to choose the appropriate agent and dosage in veterinary antimicrobial treatment. Also, some antibiotics are critical in human medicine, while some can only be used in veterinary practice (e.g. enrofloxacin, marbofloxacin, tiamulin and valnemulin) due to toxicity



**Fig. 2.** The annotated Manhattan plot of the pyseer GWAS results in case of Tylvalosin. The results are based on the analyses of 105 *Mycoplasma anserisalpinitidis* strains, excluding the atypical *M. anserisalpinitidis* strains. The X axis denotes the position of the k-mers in the type strains (ATCC BAA-2147) genome. The Y axis denotes the  $-\log_{10}$  transformation of the  $p$ -values (e.g.  $-\log_{10}(10^{-8}) = 8$ ). The annotations are based on the CDS the k-mer was mapped to. Red color denotes CDSs coding hypothetical proteins, while green color shows proteins with known function or domain. The annotations show the CDSs where k-mers with significant associations were mapped. The k-mers were annotated based on protein function and whether or not it can play a role in antibiotic resistance. Certain k-mers have been annotated which may not be present in a protein with known role in antibiotic resistance, nevertheless these CDSs are of interest because of their unusualness or their potential role in the spread of antibiotic resistance (in tylvalosin: chromate transporter, *uvrC*, YaaA family protein, MurP/RpiR family transcriptional regulator, ribonuclease J and *ileS*; in lincomycin: *tsf*, chromate transporter, *uvrC*, SmR/MutS family protein, ZIP family metal transporter, *mmnE*; in spectinomycin: *dnaB*, transglutaminase domain-containing protein, *truB*).



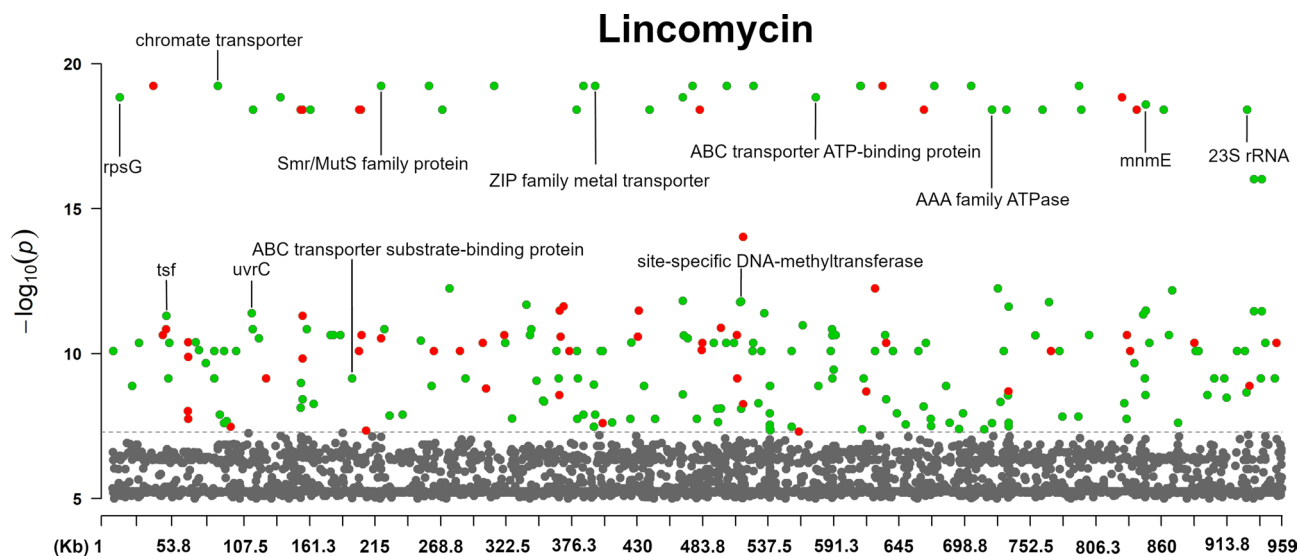


**Fig. 4.** The annotated Manhattan plot of the pyseer GWAS results in case of Enrofloxacin. The results are based on the analyses of 105 *Mycoplasma anserisalpingtonis* strains, excluding the atypical *M. anserisalpingtonis* strains. The X axis denotes the position of the k-mers in the type strains (ATCC BAA-2147) genome. The Y axis denotes the  $-\log_{10}$  transformation of the  $p$ -values (e.g.  $-\log_{10}(10^{-8})=8$ ). The annotations are based on the CDS the k-mer was mapped to. Red color denotes CDSs coding hypothetical proteins, while green color shows proteins with known function or domain. The annotations show the CDSs where k-mers with significant associations were mapped. The k-mers were annotated based on protein function and whether or not it can play a role in antibiotic resistance. Certain k-mers have been annotated which may not be present in a protein with known role in antibiotic resistance, nevertheless these CDSs are of interest because of their unusualness or their potential role in the spread of antibiotic resistance (in tylvalosin: chromate transporter, *uvrC*, YaaA family protein, MurP/RpiR family transcriptional regulator, ribonuclease J and *ileS*; in lincomycin: *tsf*, chromate transporter, *uvrC*, SmR/MutS family protein, ZIP family metal transporter, *mmmE*; in spectinomycin: *dnaB*, transglutaminase domain-containing protein, *truB*).

Fluoroquinolones (enrofloxacin) inhibit DNA replication by targeting DNA gyrase and topoisomerase IV enzymes, while bacteria can evolve resistance by mutations in the coding genes of these proteins along with efflux of the agents. We found multiple k-mers in CDSs that corresponded with DNA repair (like *recA*, *uvrB* and *ruvA*) or DNA replication (like *dnaN* or *dnaG*), along with various efflux pump component genes.

Aminoglycosides interact with the 30S ribosomal subunit of 16S rRNA causing misreading and truncated proteins and cell death. The resistance mechanism of the bacteria is the modification of aminoglycosides through various enzymes, like acetyltransferases, phosphotransferases, nucleotidyltransferases or 16S ribosomal methylases; however, the mutation of the 16S rRNA gene and increased efflux can also contribute to elevated resistance. We found a diverse set of CDSs with significant k-mers, like a phosphoribosyl transferase, methyltransferases and efflux pump subunits. Interestingly, significant k-mer hits were detected in the DNA gyrase subunit B and DNA topoisomerase IV subunit A CDSs associated with spectinomycin resistance. These two enzyme subunits are not known to be involved in the aminoglycoside resistance, and further studies are required to validate their putative role. In silico methylation screening or RNA sequencing could be employed to further elucidate their functions. Alternatively, repeating GWA study with a larger and more diverse population set of samples might serve to corroborate or refute the current finding.

Our study found that various transferases, especially methyltransferases often show a significant association with altered susceptibility to multiple antimicrobial agents. This corresponds to the fact that various transferases can play key roles in the resistance mechanisms of multiple antibiotics<sup>36</sup>. The second most common type of CDSs with significant hits were either members of efflux pumps or had efflux pump activity. Increase of efflux activity is an efficient way for a bacterium to defend itself against multiple antimicrobial agents. It is often accompanied by a decreased influx of the antibiotic, and these two mechanisms in tandem can lower the antimicrobial agent inside a cell<sup>39</sup>. Significant associations of members of efflux pumps or with efflux pump activity in the present study correspond to the results described earlier, identifying genetic markers in efflux pumps (like ABC-transporter ATP-binding proteins, ABC transporter permease subunits or MATE family efflux transporter) in *M. anserisalpingtonis* mutants with lower susceptibility for certain antimicrobial agents<sup>14</sup>. The two types of CDSs that were the third most common types with significant hits were ligases, many of which are transfer RNA or DNA ligases; and CDSs that code proteins taking part in DNA repair, like recombinases or excinucleases. Furthermore, multiple other types of CDSs previously associated with susceptibility to various antimicrobial agents were detected also in the present study, such as small and large ribosomal subunit protein, 23S rRNA, or gyrases and isomerase genes. However, there were multiple CDSs with significant k-mers, that have not been directly linked with any specific antibiotic resistance mechanisms, but most likely participate in multiple resistance mechanisms. One such group of CDSs was part of the quorum sensing pathway in *M.*



**Fig. 5.** The annotated Manhattan plot of the pyseer GWAS results in case of Lincomycin. The results are based on the analyses of 105 *Mycoplasma anserisalpinitidis* strains, excluding the atypical *M. anserisalpinitidis* strains. The X axis denotes the position of the k-mers in the type strains (ATCC BAA-2147) genome. The Y axis denotes the  $-\log_{10}$  transformation of the  $p$ -values (e.g.  $-\log_{10}(10^{-8}) = 8$ ). The annotations are based on the CDS the k-mer was mapped to. Red color denotes CDSs coding hypothetical proteins, while green color shows proteins with known function or domain. The annotations show the CDSs where k-mers with significant associations were mapped. The k-mers were annotated based on protein function and whether or not it can play a role in antibiotic resistance. Certain k-mers have been annotated which may not be present in a protein with known role in antibiotic resistance, nevertheless these CDSs are of interest because of their unusualness or their potential role in the spread of antibiotic resistance (in tylosin: chromate transporter, *uvrC*, YaaA family protein, MurP/RpiR family transcriptional regulator, ribonuclease J and *ileS*; in lincomycin: *tsf*, chromate transporter, *uvrC*, SmR/MutS family protein, ZIP family metal transporter, *mnmE*; in spectinomycin: *dnaB*, transglutaminase domain-containing protein, *truB*).

*anserisalpinitidis* based on the KEGG database. Another important finding was the detection of CDSs that have been described in a previous study<sup>32</sup> as part of prophage-like sequences. This corresponds to the fact that phages can integrate into a bacterial genome under appropriate circumstances and participate in the survival of the bacteria. The presence of multiple hypothetical proteins and uncharacterized CDSs possibly points toward multiple unknown antibiotic resistance pathways. Another hypothesis is that these CDSs are so far non-characterized members of established pathways.

The identification of the k-mers associated with antibiotic susceptibility promotes the development of novel diagnostic tests and support the understanding of resistance mechanisms not only in *M. anserisalpinitidis*, but supposedly in other mycoplasmas and bacteria species. In addition, it is most likely that the combined effects of antibiotic resistance mechanisms maximize the survival of the pathogens.

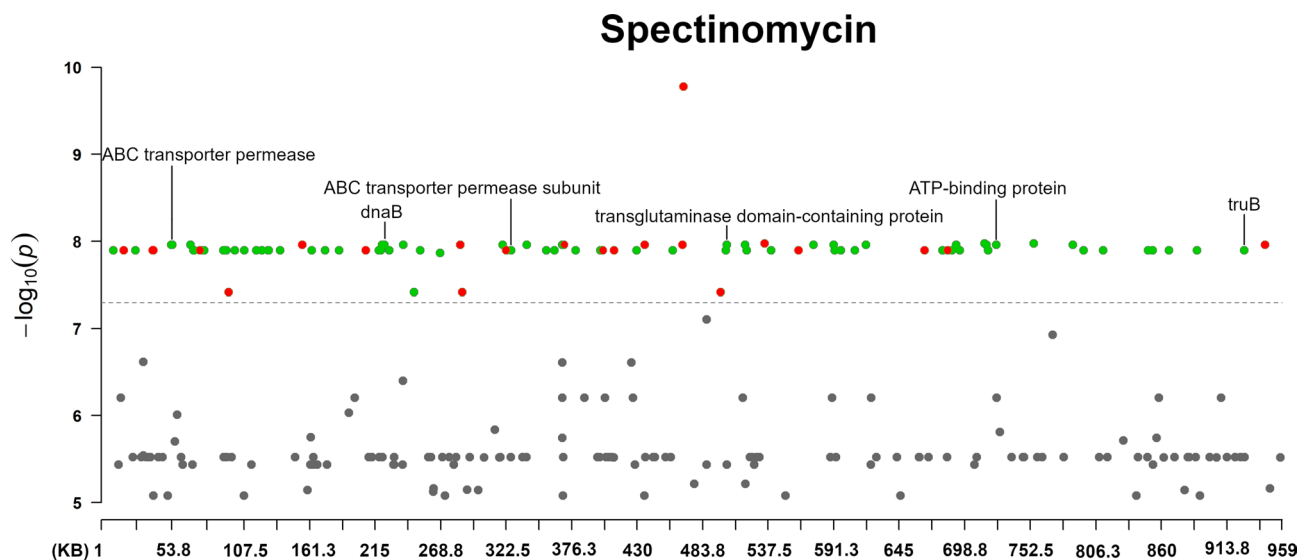
## Conclusions

While the results of our study correlated with most described antibiotic resistance mechanisms (efflux pumps, methyltransferases, mutations in target genes of antimicrobials), the identification of multiple unidentified or hypothetical protein genes suggests a more intricate genetic background for antimicrobial resistance. The detected genes associated with putative prophage-like sequences suggest potential horizontal gene transfer events that could facilitate the acquisition of novel resistance mechanisms. The genetic background of antimicrobial resistance in the case of *M. anserisalpinitidis* is composed of multiple factors, requiring further investigations to clarify their precise roles.

## Materials and methods

### Antimicrobial susceptibility testing

The antimicrobial susceptibility profiles of 110 *M. anserisalpinitidis* strains have been analyzed in this study. The strains originated from domestic and swan geese and a domestic duck between 1983 and 2019, and were collected from Hungary ( $n = 99$ ), Poland ( $n = 8$ ), China ( $n = 2$ ), and Vietnam ( $n = 1$ ), from the cloaca ( $n = 59$ ), phallus and phallus lymph ( $n = 36$ ), trachea ( $n = 5$ ), follicle ( $n = 5$ ), oviduct ( $n = 1$ ), semen ( $n = 3$ ) or the lung and air sac ( $n = 1$ ) of the animals. The metadata of the strains can be found in Supplementary Table 1. According to the written declaration (reference number: VMRI/2018/0024) of the Ethics Committee of the HUN-REN Veterinary Medical Research Institute ethical approval was not required for the study as the samples were taken during routine diagnostic examinations with the written consent of the owner.



**Fig. 6.** The annotated Manhattan plot of the pyseer GWAS results in case of Spectinomycin. The results are based on the analyses of 105 *Mycoplasma anserisalpingtonis* strains, excluding the atypical *M. anserisalpingtonis* strains. The X axis denotes the position of the k-mers in the type strains (ATCC BAA-2147) genome. The Y axis denotes the  $-\log_{10}$  transformation of the  $p$ -values (e.g.  $-\log_{10}(10^{-8}) = 8$ ). The annotations are based on the CDS the k-mer was mapped to. Red color denotes CDSs coding hypothetical proteins, while green color shows proteins with known function or domain. The annotations show the CDSs where k-mers with significant associations were mapped. The k-mers were annotated based on protein function and whether or not it can play a role in antibiotic resistance. Certain k-mers have been annotated which may not be present in a protein with known role in antibiotic resistance, nevertheless these CDSs are of interest because of their unusualness or their potential role in the spread of antibiotic resistance (in tylvalosin: chromate transporter, *uvrC*, YaaA family protein, MurP/RpiR family transcriptional regulator, ribonuclease J and *ileS*; in lincomycin: *tsf*, chromate transporter, *uvrC*, SmR/MutS family protein, ZIP family metal transporter, *mmE*; in spectinomycin: *dnaB*, transglutaminase domain-containing protein, *truB*).

The isolation and cell propagation necessary for the antimicrobial susceptibility testing were performed as described in previous studies<sup>12,31,33,40</sup>. In brief, the isolation and propagation were accomplished in Oxoid Mycoplasma broth medium (pH 7.8) (Thermo Fisher Scientific, Inc., Waltham, MA, USA) supplemented with 0.5% (wt/vol) sodium pyruvate, 0.5% (wt/vol) glucose, 0.15% l-arginine hydrochloride and 0.005% (wt/vol) phenol red, incubating the cultures at 37 °C for 1–2 days. Nine antimicrobial agents were tested using the broth microdilution method: a pleuromutilin (tiamulin); two tetracyclines (oxytetracycline and doxycycline); three macrolides (tylosin, tylvalosin, tilmicosin); a lincosamide (lincomycin); an aminocyclitol (spectinomycin) and a fluoroquinolone (enrofloxacin). The dilution and storage of the antibiotics, broth microdilution examinations, and determination of the MIC values were performed as established previously<sup>10</sup>.

In this study the MIC values were arranged in an orderly array and the middle value ( $0.5 \times n$  (the number of strains—110 in this study)) was selected for MIC<sub>50</sub>. The MIC<sub>90</sub> value was calculated similarly; however, in this case the 99<sup>th</sup> value was selected ( $0.9 \times n$ ). The antibiotics were applied two separate ranges, for tylosin, tilmicosin, tylvalosin, lincomycin, oxytetracycline, spectinomycin of 0.25–64 µg/mL, while for doxycycline, enrofloxacin and tiamulin the range was 0.039–10 µg/mL. The MIC values that exceeded the range of the assay were included in the study with a 1.5 multiplier, denoting the fact these values were higher than the range, but could not be determined exactly. Out of the 110 *M. anserisalpingtonis* strains, the MIC values for 31 isolates were previously determined, whereas susceptibility testing of 79 isolates was either performed or redone during this study using the same method<sup>12,40</sup>. The *M. anserisalpingtonis* ATCC BAA-2147 type strain was tested on each microtiter plate to confirm the validity of the results.

### Whole-genome sequencing, assembly, and annotation

The *M. anserisalpingtonis* whole-genome sequences originated from previous studies<sup>31,32</sup>. To summarize, the 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. 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.).

The quality check of the short reads was performed using FastQC software version 0.11.9<sup>41</sup> the quality threshold was a Phred score of 30. A draft genome was assembled for all of the strains with SPAdes version 3.13.1<sup>42</sup>. The contigs of the draft genomes were ordered using ATCC BAA-2147 with the MAUVE algorithm version 20150226<sup>43</sup> and annotated with the NCBI Prokaryote Genome Annotation Pipeline (PGAP) version 6.2<sup>44</sup>.

	Enrofloxacin	Tetracycline	Spectinomycin	Macrolides	Lincomycin	TOTAL
<b>Unknown function</b>						
Hypothetical protein	0	1	0	2	1	4
Unknown function	6	0	15	19	24	64
<b>Enzymatic activity</b>						
Aminotransferase	0	0	0	1	1	2
ATP synthase activity	0	0	0	1	1	2
Catalytic activity	0	0	0	0	1	1
Hydrolase	0	0	1	1	3	5
Isomerase	1	0	3	0	1	5
Kinase	1	0	0	2	1	4
Ligase	1	1	2	4	4	12
Nuclease	1	0	3	4	2	10
Protease	1	0	2	1	0	4
Topoisomerase	0	0	3	1	2	6
Transferase	3	0	6	4	10	23
<b>DNA Related Functions</b>						
26S rRNA genes	0	0	0	1	1	2
DNA repair	3	0	3	4	3	13
DNA replication	0	0	2	2	3	7
Intergenic region	0	0	1	1	1	3
Promoter	1	0	0	1	1	3
Repeat region	1	0	0	0	0	1
Transposase	0	0	1	2	2	5
<b>Ribosomal Functions</b>						
Large ribosomal subunit	0	0	0	2	2	4
Small ribosomal subunit	2	0	1	2	2	7
Protein biosynthesis	1	0	0	1	2	4
Ribosome biogenesis	1	0	1	1	2	5
<b>Transcription and Translation Regulation</b>						
Elongation factor	0	0	0	1	1	2
RNA polymerase	0	0	1	1	0	2
Signal processing	0	0	0	1	1	2
Translocase	1	0	0	0	2	3
Transcription regulation	2	0	0	2	1	5
<b>Metabolic Pathways</b>						
Amino sugar and nucleotide sugar metabolism	0	0	0	0	1	1
Carbon metabolism	0	0	0	1	0	1
Chromosome partitioning proteins	1	0	1	0	0	2
Glycolysis / Gluconeogenesis	0	0	0	0	1	1
Oxireductase	1	0	3	2	2	8
Pyruvate metabolism	0	0	0	1	0	1
Transfer RNA biogenesis	0	0	0	1	0	1
tRNA processing	0	0	0	0	1	1
<b>Transport and Defense</b>						
Antiviral defense	0	0	0	1	1	2
Chaperone	0	0	0	1	1	2
Efflux pump	0	0	5	8	8	21
Transmembrane transport	0	0	0	1	0	1
Transporter	0	0	0	3	4	7

**Table 1.** The CDSs with significant k-mer hits.

The CDSs are grouped by functions (rows) and the number of CDSs with significant k-mers in the group are given. The color coding ranges from lowest (red), medium (yellow) to the highest (green) values.

## Genotype: phenotype association test and promoter mapping

The preparatory steps necessary for the study, the estimation of genomic distances and counting k-mers were performed based on the pyseer documentation<sup>45</sup>. The core genome multi-locus sequence typing (cgMLST) schema from previous studies<sup>19,27</sup> was used in this study. The genetic distance of the strains was calculated from the cgMLST schema with the script provided by the pyseer software in order to correct for the population structure. In the following step, the k-mers were counted using fsm-lite software version 1.0<sup>46</sup>, as per the best practice in the pyseer documentation. The number dimensions to retain was calculated using the output of a scree plot python script.

The genome wide association study was conducted with the pyseer software version 1.3.11<sup>47</sup>, following the tutorial and best practices on the pyseer software's documentation. In short, we used three separate files: MIC values of the samples, the distance matrix from the cgMLST schema, and the k-mers counted by fsm-lite in the previous step. As there are no official cutoff for antibiotic susceptibility in the case of *M. anserisalpinitidis*, the MIC values were treated as continuous variable. Linear mixed model was used in the study to better account for the relatedness of the strains. The association study was repeated for each antimicrobial agent, nine times in total. Due to multiple statistical testing, a more stringent *p*-value criterion is used in genome wide studies, usually  $p < 5 \times 10^{-8}$ . Statistically significant k-mers were mapped to the type strain (ATCC BAA-2147; GenBank Accession number: CP042295)<sup>33</sup>. Promotech version 1.0<sup>48</sup> software was used to search for promoter regions in four strains with complete whole genomes (ATCC BAA-2147; MYCAV93, GenBank Accession number: CP041663; MYCAV177, GenBank Accession number: CP041664 and MYCAV271, GenBank Accession number: CP083178). The promoter regions with a score of 0.8 (as there are only few promoters described in *Mycoplasmas*, this stricter than usual criteria was chosen) have been used to annotate the type strain in Geneious Prime 2023.1.1<sup>49</sup>. The significant k-mers and promoter regions were cross-referenced. After running the association study significant background noise could be observed in the Manhattan plots, as such the five atypical strains (MYCAV61, MYCAV93, MYCAV783, MYCAV785 and MYCAV903), that showed significant divergence in genomic composition, were omitted and the whole procedure was run again, with new distance matrix, k-mers and phenotype files.

## Data availability

The short reads can be accessed in the NCBI SRA database: PRJNA602215 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA602215>); PRJNA856806 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA856806>); PRJNA650261 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA650261>); PRJNA602206 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA602206>). The draft genome assemblies and the CDSs with significant k-mer hit can be accessed at: under: [https://figshare.com/articles/dataset/\\_i\\_Mycoplasma\\_anserisalpinitidis\\_i\\_genome\\_wide\\_association\\_study\\_data/25378600](https://figshare.com/articles/dataset/_i_Mycoplasma_anserisalpinitidis_i_genome_wide_association_study_data/25378600).

Received: 16 June 2024; Accepted: 6 February 2026

Published online: 24 February 2026

## References

- Volokhov, D. V. et al. *Mycoplasma anserisalpinitidis* sp. nov., isolated from European domestic geese (*Anser anser domesticus*) with reproductive pathology. *Int. J. Syst. Evol. Microbiol.* <https://doi.org/10.1099/ijsem.0.004052> (2020).
- Stipkovits, L., Varga, Z., Dobos-Kovacs, M. & Santha, M. Biochemical and serological examination of some *Mycoplasma* strains of goose origin. *Acta Vet. Hung.* **32**, 117–125 (1984).
- Grözner, D. et al. Multilocus sequence typing of the goose pathogen *Mycoplasma anserisalpinitidis*. *Vet. Microbiol.* <https://doi.org/10.1016/j.vetmic.2020.108972> (2021).
- Gyuranecz, M. et al. Isolation of *Mycoplasma anserisalpinitidis* from swan goose (*Anser cygnoides*) in China. *BMC Vet. Res.* **16**, 178. <https://doi.org/10.1186/s12917-020-02393-5> (2020).
- Carnaccini, S. et al. A novel *Mycoplasma* sp. associated with phallus disease in goose breeders: Pathological and bacteriological findings. *Avian Dis.* **60**, 437–443 (2016).
- Sprygin AV, Volokhov DV, Irza VN, Drygin VV. Detection and genetic identification of *Mycoplasma* sp. 1220 in geese in the Russian Federation and Ukraine. *Agric Biol.* 2012;:87–95. <https://doi.org/10.15389/AGROBIOLOGY.2012.2.87ENG>.
- Dobos-Kovacs, M., Varga, Z., Czifra, G. & Stipkovits, L. Salpingitis in geese associated with *Mycoplasma* sp. strain 1220. *Avian Pathol.* **38**, 239–243. <https://doi.org/10.1080/03079450902912127> (2009).
- Stipkovits, L. & Kempf, I. Mycoplasmoses in poultry. *Rev. Sci. Tech.* **15**, 1495–1525 (1996).
- Stipkovits, L. & Szathmary, S. *Mycoplasma* infection of ducks and geese. *Poult. Sci.* **91**, 2812–2819. <https://doi.org/10.3382/ps.2012-02310> (2012).
- Hannan, P. C. T. Guidelines and recommendations for antimicrobial minimum inhibitory concentration (MIC) testing against veterinary mycoplasma species. *Vet. Res.* **31**, 373–395. <https://doi.org/10.1051/vetres:2000100> (2000).
- Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J. Anaesthesiol. Clin. Pharmacol.* **33**, 300–305. [https://doi.org/10.4103/joacp.JOACP\\_349\\_15](https://doi.org/10.4103/joacp.JOACP_349_15) (2017).
- Grözner, D. et al. Identification and detection of mutations potentially associated with decreased susceptibility to macrolides and lincosamin in *Mycoplasma anserisalpinitidis* isolates. *Vet. Microbiol.* <https://doi.org/10.1016/j.vetmic.2022.109362> (2022).
- Beko, K., Nagy, E. Z., Grözner, D., Kreizinger, Z. & Gyuranecz, M. Biofilm formation and its impact on environmental survival and antibiotic resistance of *Mycoplasma anserisalpinitidis* strains. *Acta Vet. Hung.* **70**, 184–191. <https://doi.org/10.1556/004.2022.00029> (2022).
- Nagy, E. Z. et al. Phenotypic and genetic insights into efflux pump mechanism in *Mycoplasma anserisalpinitidis*. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2023.1216893> (2023).
- Ozaki, K. et al. Functional SNPs in the lymphotoxin- $\alpha$  gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654. <https://doi.org/10.1038/ng1047> (2002).
- Haines, J. L. et al. Complement factor H variant increases the risk of age-related macular degeneration. *Science* **2005**(308), 419–421. [https://doi.org/10.1126/SCIENCE.1110359/SUPPL\\_FILE/HAINES.SOM.PDF](https://doi.org/10.1126/SCIENCE.1110359/SUPPL_FILE/HAINES.SOM.PDF) (1979).
- Habib, F., Johnson, A. D., Bundschuh, R. & Janies, D. Large scale genotype-phenotype correlation analysis based on phylogenetic trees. *Bioinformatics* **23**, 785–788. <https://doi.org/10.1093/bioinformatics/btm003> (2007).

18. Wozniak, M., Tiuryn, J. & Wong, L. GWAMAR: Genome-wide assessment of mutations associated with drug resistance in bacteria. *BMC Genomics* **15**(10), S10. <https://doi.org/10.1186/1471-2164-15-S10-S10> (2014).
19. Lees, J. A. et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat. Commun.* <https://doi.org/10.1038/ncomms12797> (2016).
20. Earle, S. G. et al. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat. Microbiol.* **1**, 1–21. <https://doi.org/10.1038/nmicrobiol.2016.41> (2016).
21. Jaillard, M. et al. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet.* **14**, 1–28. <https://doi.org/10.1371/journal.pgen.1007758> (2018).
22. Bokma, J. et al. Genome-wide association study reveals genetic markers for antimicrobial resistance in *Mycoplasma bovis*. *Microbiol. Spectr.* <https://doi.org/10.1128/spectrum.00262-21> (2021).
23. Hang, N. T. L. et al. Whole genome sequencing, analyses of drug resistance-conferring mutations, and correlation with transmission of *Mycobacterium tuberculosis* carrying katG-S315T in Hanoi, Vietnam. *Sci. Rep.* <https://doi.org/10.1038/s41598-019-51812-7> (2019).
24. Jiang, Y. et al. Comparative genomic sequencing to characterize *Mycoplasma pneumoniae* genome, typing, and drug resistance. *Microbiol. Spectr.* <https://doi.org/10.1128/spectrum.03615-23> (2024).
25. Malekian, N., Sainath, S., Al-Fatlawi, A. & Schroeder, M. Word-based GWAS harnesses the rich potential of genomic data for *E. coli* quinolone resistance. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2023.1276332> (2023).
26. Jaillard, M. et al. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1007758> (2018).
27. Weber, R. E. et al. Genome-wide association studies for the detection of genetic variants associated with daptomycin and ceftaroline resistance in *Staphylococcus aureus*. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2021.639660> (2021).
28. Ledger, L., Eidt, J. & Cai, H. Y. Identification of antimicrobial resistance-associated genes through whole genome sequencing of *Mycoplasma bovis* isolates with different antimicrobial resistances. *Pathogens* **9**, 1–13. <https://doi.org/10.3390/pathogens9070588> (2020).
29. Waldner, M. et al. Genome-wide association study of nucleotide variants associated with resistance to nine antimicrobials in *Mycoplasma bovis*. *Microorganisms* <https://doi.org/10.3390/microorganisms10071366> (2022).
30. Bokma, J. et al. Genome-wide association study reveals genetic markers for antimicrobial resistance in *Mycoplasma bovis*. *Microbiol. Spectr.* **9**, 1–14. <https://doi.org/10.1128/spectrum.00262-21> (2021).
31. Kovács, Á. B. et al. The core genome multi-locus sequence typing of *Mycoplasma anserisalpingitidis*. *BMC Genomics* **21**, 403. <https://doi.org/10.1186/s12864-020-06817-2> (2020).
32. Kovács, Á. B. et al. Novel prophage-like sequences in *Mycoplasma anserisalpingitidis*. *Infect. Genet. Evol.* <https://doi.org/10.1016/j.meegid.2021.104886> (2021).
33. Kovács, Á. B. et al. Characterization of atypical *Mycoplasma anserisalpingitidis* strains. *Vet. Microbiol.* <https://doi.org/10.1016/j.vetmic.2023.109722> (2023).
34. Adrian-Kalchhauser, I. et al. Understanding “non-genetic” inheritance: Insights from molecular-evolutionary crosstalk. *Trends Ecol. Evol.* **35**, 1078–1089. <https://doi.org/10.1016/j.tree.2020.08.011> (2020).
35. Peryere, S. & Tardy, F. Integrating the human and animal sides of mycoplasmas resistance to antimicrobials. *Antibiotics* **10**, 1–23. <https://doi.org/10.3390/antibiotics10101216> (2021).
36. Darby, E. M. et al. Molecular mechanisms of antibiotic resistance revisited. *Nat Rev Microbiol.* **21**, 280–295. <https://doi.org/10.1038/s41579-024-01014-4> (2023).
37. Paukner, S. & Riedl, R. Pleuromutilins: Potent drugs for resistant bugs—mode of action and resistance. *Cold Spring Harb. Perspect. Med.* <https://doi.org/10.1101/cshperspect.a027110> (2017).
38. Fernandez, N. L. et al. DNA methylation and RNA-DNA hybrids regulate the single-molecule localization of a DNA methyltransferase on the bacterial nucleoid. *MBio* <https://doi.org/10.1128/mbio.03185-22> (2023).
39. Sultan, I. et al. Antibiotics, resistance and resistance mechanisms: A bacterial perspective. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2018.02066> (2018).
40. Grözner, D. et al. 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> (2016).
41. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 3 Feb 2020.
42. Bankevich, A. et al. 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> (2012).
43. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403. <https://doi.org/10.1101/gr.2289704.tion> (2004).
44. Tatusova, T. et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624. <https://doi.org/10.1093/nar/gkw569> (2016).
45. pyseer documentation — pyseer 1.3.10 documentation. <https://pyseer.readthedocs.io/en/master/>. Accessed 21 Mar 2024.
46. Niko V, Lees JA. No Title. <https://github.com/nvalimak/fsm-lite>.
47. Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. Pyseer: A comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**, 4310–4312. <https://doi.org/10.1093/bioinformatics/bty539> (2018).
48. Chevez-Guardado, R. & Peña-Castillo, L. Promotech: A general tool for bacterial promoter recognition. *Genome Biol.* **22**, 1–16. <https://doi.org/10.1186/s13059-021-02514-9> (2021).
49. Kears, M. et al. 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> (2012).

## Acknowledgements

The authors would like to thank Veronika Hrivnák for her significant assistance in conducting the MIC testing.

## Author contributions

Áron B. Kovács—Conceptualization, Methodology, Formal analysis, Data Curation, Writing—Original Draft. Enikő Wehmann—Methodology, Data Curation, Writing—Review and Editing. Katinka Bekő Resources, Writing—Review and Editing. Dénes Grözner—Resources, Writing—Review and Editing. Krisztina Bali—Resources, Writing—Review and Editing. Zsuzsa Kreizinger—Resources, Writing—Review and Editing. Anna Sawicka—Resources, Writing—Review and Editing. Krisztián Bányai—Resources, Writing—Review and Editing. Miklós Gyuranecz—Conceptualization, Writing—Review and Editing, Supervision, Funding acquisition.

## Funding

Open access funding provided by HUN-REN Veterinary Medical Research Institute. This work was supported

by the KKP19 (129751) grant of the National Research, Development and Innovation Office, Hungary, the SA-27/2021 grant of the Eötvös Loránd Research Network, the Project no. RRF-2.3.1-21-2022-00001 which has been implemented with the support provided by the Recovery and Resilience Facility (RRF), financed under the National Recovery Fund budget estimate, RRF-2.3.1-21 funding scheme and the support provided by the Ministry of Innovation and Technology of Hungary (legal successor: Ministry of Culture and Innovation of Hungary) from the National Research, Development and Innovation Fund, financed under the TKP2021-EGA-01 funding scheme of the National Research, Development and Innovation Office. Prepared with the professional support of the Doctoral Student Scholarship Program of the Co-operative Doctoral Program of the Ministry of Innovation and Technology, financed from the National Research, Development and Innovation Fund. The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

### Ethics approval and consent to participate

According to the written declaration (Reference Number: IVMR/2019/0023) of the Ethics Committee of the HUN-REN Veterinary Medical Research Institute ethical approval was not required for the study as the samples were taken during routine diagnostic examinations with the written consent of the owner.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-39804-w>.

**Correspondence** and requests for materials should be addressed to M.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026