

1 Genome-wide association study of *Mycoplasma*

2 *anserisalpingitidis* strains for antibiotic susceptibility

3 Áron B. Kovács^{a,b}, Enikő Wehmann^{a,b}, Katinka Bekő^a, Dénes Gróznér^a, Krisztina Bali^a, Zsuzsa
4 Kreizinger^a, Anna Sawicka^c, Krisztián Bányai^{a,b,d}, Miklós Gyuranecz^{a,b,d,e*}

5
6 ^aHUN-REN Veterinary Medical Research Institute, H-1143
7 Hungária krt. 21, Budapest, Hungary

8 ^bNational Laboratory of Infectious Animal Diseases, Antimicrobial Resistance, Veterinary
9 Public Health and Food Chain Safety, H-1143 Hungária krt. 21, Budapest, Hungary

10 ^cDepartment of Poultry Diseases, National Veterinary Research Institute, Aleja Partyzantow
11 57, 24-100, Pulawy, Poland

12 ^dUniversity of Veterinary Medicine, H-1078 István u. 2, Budapest, Hungary

13 ^eMolliscience Kft., H-2051 Március 15. utca 1., Biatorbágy, Hungary

14 *Corresponding author

15 E-mail: m.gyuranecz@gmail.com

16

Abstract

Mycoplasma anserisalpingitidis 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. anserisalpingitidis* 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 finding the genetic background of antimicrobial resistance of *M. anserisalpingitidis* 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. anserisalpingitidis* 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

Introduction

Mycoplasma anserisalpingitidis is a waterfowl facultative pathogenic bacterium, primarily found in geese and occasionally infecting ducks [1]. Mycoplasmas lack cell wall and are among the smallest known prokaryotes in terms of genome size. *Mycoplasma anserisalpingitidis* was first isolated in 1983 in Hungary [2] and was fully characterized in 2020 by Volokhov et al. [1]. This pathogen has been frequently detected in Hungarian geese farms [3], and it is probably present around the whole world [4, 5], as it has been found in geese in Poland, Ukraine and the Russian Federation [6], China [4] and Vietnam. *Mycoplasma anserisalpingitidis* 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 the eggs, increased embryo lethality, phallus and cloaca inflammation, peritonitis, salpingitis, and airsacculitis, leading to high economic losses [7–9].

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. anserisalpingitidis*. In case of a disease occurrence, antimicrobial treatment is the primarily option to be used against *M. anserisalpingitidis*.

There is a known increase in resistance toward antimicrobial agents worldwide, both in human and veterinary medicine, which necessitates rigorous antimicrobial susceptibility testing for the application of targeted antimicrobial therapy. One of the commonly used methods for antimicrobial susceptibility testing of mycoplasmas is the broth microdilution method [10]; however, it can take multiple weeks for this method to be completed. Such a long period makes the diagnostic results too late to make treatment decision and can lead to significant economic losses and demands the development of rapid diagnostic tests.

Multiple environmental factors can promote the spread of resistance to antimicrobials such as 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 [11]. 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

70 methods for antibiotic susceptibility testing by designing assays targeting these genetic
71 markers. Several studies analyzed the antibiotic susceptibility in *M. anserisalpiginidis*
72 examining single nucleotide polymorphisms (SNPs), biofilm formation and efflux pump
73 mechanisms [12–14].

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

85 As GWA studies use various statistic methods (e.g. Fishers exact test, binomial test, fixed effect
86 generalized linear regression); therefore, the number of samples necessary for the experiment
87 to be successful depends on how strongly a certain phenotype in question correlates with the
88 genotype. This number can vary from approximately a hundred to thousands of samples, with
89 more samples allowing the detection of genetic variances with smaller impacts. With advances
90 in whole genome sequencing, it has become ever easier and cheaper to attain reliable whole-
91 genome sequences (WGS) of different bacteria. As mentioned prior it is important for the
92 sequencing to be precise and the coverage of the sequences to be deep and high standard
93 regarding the Phred score (indicates the measure of the base quality) and the depth (describes
94 how often a given nucleotide has been read in the experiment) of the sequencing should be
95 set up. The third important factor in GWAS is phenotypic data, which can range from binary
96 characteristics (such as manifestation of an illness in animals and humans or the ability to
97 metabolize certain molecules) to continuous (like the physical parameters of the organisms or
98 antimicrobial resistance of a microbe). Certain characteristics are relatively easy to note (e.g.
99 host species, physical parameters), while others have to be acquired through various methods
100 (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 unitigs [21].

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* [22–24]. 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 MIC₅₀ and MIC₉₀ values recorded for tilmicosin (≥ 64 $\mu\text{g/ml}$ for both MIC₅₀ and MIC₉₀). Tylosin inhibited at 16 $\mu\text{g/ml}$ the 50% of the strains and at ≥ 64 $\mu\text{g/ml}$ the 90% of the strains, while tylvalosin showed the lowest MIC₅₀ value (0.5 $\mu\text{g/ml}$) and MIC₉₀ value (2 $\mu\text{g/ml}$) against the strains. The MIC values for lincomycin clustered around the MIC₅₀ value (2 $\mu\text{g/ml}$), although one isolate yielded high MIC values (> 64 $\mu\text{g/ml}$). Tetracyclines displayed broad ranges of the MIC values (≤ 0.25 to > 64 $\mu\text{g/ml}$ for oxytetracycline and 0.078 to > 10 $\mu\text{g/ml}$ for doxycycline), with high MIC₅₀ (16 and 5 $\mu\text{g/ml}$, respectively) and MIC₉₀ values (64 and 10 $\mu\text{g/ml}$, respectively). Spectinomycin exhibited a MIC₅₀ of 8 $\mu\text{g/ml}$, with the majority of strains displaying MIC₅₀ or higher values. The MIC values for enrofloxacin showed a wide range (1.25 to > 10 $\mu\text{g/ml}$), with MIC₅₀ and a MIC₉₀ values above 10 $\mu\text{g/ml}$. Tiamulin also showed a wide range of MIC values between 0.078 and 2.5 $\mu\text{g/ml}$. The MIC₅₀ value of tiamulin was 0.625 $\mu\text{g/ml}$, while the MIC₉₀ was 2.5 $\mu\text{g/ml}$. Individual MIC values against each strain are given in Supplementary table 1, while MIC value graphs are given in Supplementary figure 1.

Whole genome sequencing and assembly

Based on the whole genome sequencing from the previous studies [3, 25–27], 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 493.9x (Minimum: 51.7x; Maximum 2268.3x). 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 Figure 1. The pyseer tool found various numbers of significant ($p\text{-value} \leq 5 \times 10^{-8}$) k-mers for 5 of the 9 antimicrobial agents. No k-mer showed significant association with tiamulin resistance neither with tetracycline (oxytetracycline or doxycycline) 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 Figures 2-6 (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 YqaJ viral recombinase family protein CDS, both of which was also found in a previous putative prophage study on *M. anserisalpingtonis* [26]. The function of the CDSs with significant k-mer associations are summarized in Table 1 and detailed in Supplementary dataset 1 (available at [10.6084/m9.figshare.25378600](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 traits and unable to account for non-genetic elements of a phenotype, like methylation pattern or small non coding RNAs [28]. 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 or availability of more potent drugs. It is important to note that mycoplasmas have an intrinsic resistance toward multiple classes of antimicrobial agents, like beta-lactams, glycopeptides and fosfomycin, as these targets the cell wall, which mycoplasmas lack. Mycoplasmas are also resistant to rifampicin due to a mutation in the *rpoB* gene, while polymyxin and sulfonamides/trimethoprim are not effective as mycoplasmas lack lipopolysaccharides and the folic acid pathway, respectively [29].

Depending on the antimicrobial agent, there are multiple ways in which bacteria can react and defend themselves against it, such as decreasing the uptake of an antimicrobial agent (e.g. against aminoglycosides), altering the target of the antimicrobial agent by mutations in the genes of members of various ribosomal subunits [30](e.g. macrolides), or increased efflux pump activity [11]. Pleuromutilins affect bacterial protein synthesis by binding to the 23S rRNA [31].

190 Tetracyclines hinder the translational process of a bacterial cell by binding to 16S rRNA of the
191 30S ribosomal subunit and by preventing tRNA binding to 30S at the A site [11]. In the present
192 study significant k-mers in tiamulin, tetracyclines or in tylosin could not be detected, which
193 might be in connection with the above-mentioned difficulties in GWAS, and promotes the
194 search for novel techniques to reveal the mechanisms in these cases. Also, to account for the
195 effect of epigenetic regulation further studies are warranted, with various methods like RNA
196 sequencing or DNA methylation screening.

197 Macrolides disturb the translation of proteins by targeting the 23S rRNA and resistance
198 mechanisms also overlap with lincosamides. These include methylation of the 23S rRNA
199 through methyltransferases, mutations in the ribosome, through efflux or with macrolide
200 phosphotransferases and esterases. In 2022 our research team [12] designed mismatch
201 amplification mutation assays to detect nucleotide substitutions potentially associated with
202 decreased susceptibility to macrolides and lincomycin. They analyzed the genes encoding 50S
203 ribosomal proteins and targeted mutations in the 23S rRNA coding genes (*rrl1* and *rrl2*) along
204 with the 50S ribosomal L22 protein (*rplV*) coding genes [12]. In 2023 Nagy and co-workers [14]
205 identified genetic markers in efflux pumps (like ABC-transporter ATP-binding proteins, ABC
206 transporter permease subunits or MATE family efflux transporter) in *M. anserisalpinitidis*
207 mutants with lower susceptibility for the following antimicrobial agents: enrofloxacin,
208 tiamulin, lincomycin, tilmicosin, tylosin, tylvalosin. All of these corresponds with our findings
209 in this study regarding both macrolides (tilmicosin and tylvalosin specifically) and lincomycin.
210 Besides these, in lincomycin we found significant k-mers in other CDSs that match various
211 resistance mechanisms, like methyltransferases and further genes that code members of
212 efflux pumps. Bacterial DNA methylation has been linked to numerous fundamental
213 processes, including transcription, DNA replication, and restriction modification [32].

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

219 Aminoglycosides interact with the 30S ribosomal subunit of 16S rRNA causing misreading and
220 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 [30]. 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 [33]. 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. anserisalpinitidis* mutants with lower susceptibility for certain antimicrobial agents [14]. 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. anserisalpingitidis* based on the KEGG database. Another important finding was the detection of CDSs that have been described in a previous study [26] 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. anserisalpingitidis*, 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. anserisalpingitidis* 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. anserisalpingitidis* 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.

The isolation and cell propagation necessary for the antimicrobial susceptibility testing were performed as described in previous studies [12, 25, 27, 34]. 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 according to the recommendations of Hannan (2000) [10].

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 MIC₅₀. The MIC₉₀ value was calculated similarly; however, in this case the 99th 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 [12, 34]. 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 [25, 26]. 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 [35]. A draft genome was assembled for all of the strains with SPAdes version 3.13.1 [36]. The contigs of the draft genomes were ordered using ATCC BAA-2147 with the MAUVE algorithm version 20150226 [37] and annotated with the NCBI Prokaryote Genome Annotation Pipeline (PGAP) version 6.2 [38].

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 [39]. The core genome multi-locus sequence typing (cgMLST) schema from previous studies [19, 27] 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 [40]. 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 [41], 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. anserisalpingtonis*, 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) [27]. Promotech version 1.0 [42] 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 [43]. 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.

References

1. Volokhov DV, Gróznér D, Gyuranecz M, Ferguson-Noel N, Gao Y, Bradbury JM, et al. *Mycoplasma anserisalpingitidis* sp. nov., isolated from European domestic geese (*Anser anser domesticus*) with reproductive pathology. *Int J Syst Evol Microbiol*. 2020.
2. Stipkovits L, Varga Z, Dobos-Kovacs M, Santha M. Biochemical and serological examination of some *Mycoplasma* strains of goose origin. *Acta Vet Hung*. 1984;32:117–25.
3. Gróznér D, Kovács ÁB, Wehmann E, Kreizinger Z, Bekő K, Mitter A, et al. Multilocus sequence typing of the goose pathogen *Mycoplasma anserisalpingitidis*. *Vet Microbiol*. 2021;254 August 2020.
4. Gyuranecz M, Mitter A, Kovács ÁB, Gróznér D, Kreizinger Z, Bali K, et al. Isolation of *Mycoplasma anserisalpingitidis* from swan goose (*Anser cygnoides*) in China. *BMC Vet Res*. 2020;16:178.
5. Carnaccini S, Ferguson-Noel NM, Chin RP, Santoro T, Black P, Bland M, et al. A Novel *Mycoplasma* sp. Associated with Phallus Disease in Goose Breeders : Pathological and Bacteriological Findings A Novel *Mycoplasma* sp . Associated with Phallus Disease in Goose Breeders : Pathological and Bacteriological Findings. *Avian Dis*. 2016;60:437–43.
6. 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. doi:10.15389/AGROBIOLOGY.2012.2.87ENG.
7. Dobos-Kovacs M, Varga Z, Czifra G, Stipkovits L. Salpingitis in geese associated with *Mycoplasma* sp. strain 1220. *Avian Pathol*. 2009;38:239–43.
8. Stipkovits L, Kempf I. Mycoplasmoses in poultry. *Rev Sci Tech*. 1996;15:1495–525.
9. Stipkovits L, Szathmary S. *Mycoplasma* infection of ducks and geese. *Poult Sci*. 2012;91:2812–9.
10. Hannan PCT. Guidelines and recommendations for antimicrobial minimum inhibitory concentration (MIC) testing against veterinary mycoplasma species. *Vet Res*. 2000;31:373–95.
11. Kapoor G, Saigal S, Elongavan A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J Anaesthesiol Clin Pharmacol*. 2017;33:300–5.

377 12. Gróznér D, Bekő K, Kovács ÁB, Mitter A, Hrivnák V, Sawicka A, et al. Identification and
378 detection of mutations potentially associated with decreased susceptibility to macrolides and
379 lincomycin in *Mycoplasma anserisalpingitidis* isolates. *Vet Microbiol.* 2022;266.

380 13. Beko K, Nagy EZ, Gróznér D, Kreizinger Z, Gyuranecz M. Biofilm formation and its impact
381 on environmental survival and antibiotic resistance of *Mycoplasma anserisalpingitidis* strains.
382 *Acta Vet Hung.* 2022;70:184–91.

383 14. Nagy EZ, Kovács ÁB, Wehmann E, Bekő K, Földi D, Bányai K, et al. Phenotypic and genetic
384 insights into efflux pump mechanism in *Mycoplasma anserisalpingitidis*. *Front Microbiol.*
385 2023;14 July.

386 15. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, et al. Functional SNPs in the
387 lymphotoxin- α gene that are associated with susceptibility to myocardial infarction. *Nat*
388 *Genet.* 2002;32:650–4. doi:10.1038/ng1047.

389 16. Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P, et al. Complement factor
390 H variant increases the risk of age-related macular degeneration. *Science* (80-).
391 2005;308:419–21. doi:10.1126/SCIENCE.1110359/SUPPL_FILE/HAINES.SOM.PDF.

392 17. Habib F, Johnson AD, Bundschuh R, Janies D. Large scale genotype-phenotype correlation
393 analysis based on phylogenetic trees. *Bioinformatics.* 2007;23:785–8.

394 18. Wozniak M, Tiuryn J, Wong L. GWAMAR: Genome-wide assessment of mutations
395 associated with drug resistance in bacteria. *BMC Genomics.* 2014;15 Suppl 10:S10.
396 doi:10.1186/1471-2164-15-S10-S10.

397 19. Lees JA, Vehkala M, Välimäki N, Harris SR, Chewapreecha C, Croucher NJ, et al. Sequence
398 element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat*
399 *Commun.* 2016;7.

400 20. Earle SG, Wu CH, Charlesworth J, Stoesser N, Gordon NC, Walker TM, et al. Identifying
401 lineage effects when controlling for population structure improves power in bacterial
402 association studies. *Nat Microbiol.* 2016;1:1–21.

403 21. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, et al. A fast and agnostic
404 method for bacterial genome-wide association studies: Bridging the gap between k-mers and
405 genetic events. *PLoS Genet.* 2018;14:1–28.

406 22. Ledger L, Eidt J, Cai HY. Identification of antimicrobial resistance-associated genes through
407 whole genome sequencing of *mycoplasma bovis* isolates with different antimicrobial
408 resistances. *Pathogens.* 2020;9:1–13.

409 23. Waldner M, Kinnear A, Yacoub E, McAllister T, Register K, Li C, et al. Genome-Wide
410 Association Study of Nucleotide Variants Associated with Resistance to Nine Antimicrobials in
411 *Mycoplasma bovis*. *Microorganisms.* 2022;10.

412 24. Bokma J, Vereecke N, Nauwynck H, Haesebrouck F, Theuns S, Pardon B, et al. Genome-
413 Wide Association Study Reveals Genetic Markers for Antimicrobial Resistance in *Mycoplasma*
414 *bovis*. *Microbiol Spectr.* 2021;9:1–14.

415 25. Kovács ÁB, Kreizinger Z, Forró B, Gróznér D, Mitter A, Marton S, et al. The core genome
416 multi-locus sequence typing of *Mycoplasma anserisalpingitidis*. *BMC Genomics.* 2020;21:403.

doi:10.1186/s12864-020-06817-2.

26. Kovács ÁB, Wehmann E, Sváb D, Bekő K, Gróznér D, Mitter A, et al. Novel prophage-like sequences in *Mycoplasma anserisalpinitidis*. *Infect Genet Evol.* 2021;92.

27. Kovács ÁB, Wehmann E, Gróznér D, Bali K, Nemeshházi E, Hrivnák V, et al. Characterization of atypical *Mycoplasma anserisalpinitidis* strains. *Vet Microbiol.* 2023;280 February.

28. Adrian-Kalchhauser I, Sultan SE, Shama LNS, Spence-Jones H, Tiso S, Keller Valsecchi CI, et al. Understanding “Non-genetic” Inheritance: Insights from Molecular-Evolutionary Crosstalk. *Trends Ecol Evol.* 2020;35:1078–89. doi:10.1016/j.tree.2020.08.011.

29. Pereyre S, Tardy F. Integrating the human and animal sides of mycoplasmas resistance to antimicrobials. *Antibiotics.* 2021;10:1–23.

30. Darby EM, Trampari E, Siasat P, Gaya MS, Alav I, Webber MA, et al. Molecular mechanisms of antibiotic resistance revisited. *Nat Rev Microbiol.* 2023;21:280–95.

31. Paukner S, Riedl R. Pleuromutilins: Potent drugs for resistant bugs-mode of action and resistance. *Cold Spring Harb Perspect Med.* 2017;7.

32. Fernandez NL, Chen Z, Fuller DEH, van Gijtenbeek LA, Nye TM, Biteen JS, et al. DNA Methylation and RNA-DNA Hybrids Regulate the Single- Molecule Localization of a DNA Methyltransferase on the Bacterial Nucleoid. *MBio.* 2023;14.

33. Sultan I, Rahman S, Jan AT, Siddiqui MT, Mondal AH, Haq QMR. Antibiotics, resistome and resistance mechanisms: A bacterial perspective. *Front Microbiol.* 2018;9 SEP.

34. Gróznér D, Kreizinger Z, Sulyok KM, Rónai Z, Hrivnák V, Turcsányi I, et al. Antibiotic susceptibility profiles of *Mycoplasma* sp. 1220 strains isolated from geese in Hungary. *BMC Vet Res.* 2016;12:1–9. doi:10.1186/s12917-016-0799-0.

35. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 3 Feb 2020.

36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77.

37. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements.Pdf. *Genome Res.* 2004;14:1394–403.

38. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44:6614–24.

39. pyseer documentation — pyseer 1.3.10 documentation. <https://pyseer.readthedocs.io/en/master/>. Accessed 21 Mar 2024.

40. Niko V, Lees JA. No Title. <https://github.com/nvalimak/fsm-lite>.

41. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. pyseer: A comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics.* 2018;34:4310–2.

42. Chevez-Guardado R, Peña-Castillo L. Promotech: a general tool for bacterial promoter recognition. *Genome Biol.* 2021;22:1–16.

43. Kears e M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.

Acknowledgment

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

Authors' contributions

Áron B. Kovács - Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft

Enikő Wehmann - Methodology, Data Curation, Writing - Review & Editing

Katinka Bekő - Resources, Writing - Review & Editing

Dénes Gróznér - Resources, Writing - Review & Editing

Krisztina Bali - Resources, Writing - Review & Editing

Zsuzsa Kreizinger - Resources, Writing - Review & Editing

Anna Sawicka - Resources, Writing – Review & Editing

Krisztián Bányai - Resources, Writing - Review & Editing

Miklós Gyuranecz - Conceptualization, Writing - Review & Editing, Supervision, Funding acquisition

Data availability statement

The short reads can be accessed in the NCBI SRA database: PRJNA602215; PRJNA856806; PRJNA650261; PRJNA602206.

The draft genomes and the CDSs with significant k-mer hit can be accessed under: 10.6084/m9.figshare.25378600

481

482 **Ethics approval and consent to participate**

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

487

488 **Competing interests**

489 The authors declare that they have no competing interests.

490

491 **Funding**

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

505

506 **Figures**

507 **Figure 1.** The cgMLST phylogenetic tree of all of the studied *M. anserisalpingitidis* strains along
508 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

Figure 2 – 6. The annotated Manhattan plot of the pyseer GWAS results in case of Tylvalosin (Figure 2), Tilmicosin (Figure 3), Enrofloxacin (Figure 4), Lincomycin (Figure 5) and Spectinomycin (Figure 6). The results are based on the analyses of 105 *Mycoplasma anserisalpingitidis* strains, excluding the atypical *M. anserisalpingitidis* 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, mnxE; in spectinomycin: dnaB, transglutaminase domain-containing protein, truB).

531 **Tables and supplementary material**

532 **Table 1.** The CDSs with significant k-mer hits. The CDSs are grouped by functions (rows) and
 533 the number of CDSs with significant k-mers in the group are given.

	Enrofloxacin	Tetracycline	Spectinomycin	Macrolides	Lincomycin	TOTAL
Unknown function						
Hypothetical protein	0	1	0	2	1	4
Unknown function	6	0	15	19	24	64
Enzymatic activity						
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
DNA Related Functions						
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
Ribosomal Functions						
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
Transcription and Translation Regulation						
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
Metabolic Pathways						
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
Transport and Defense						
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

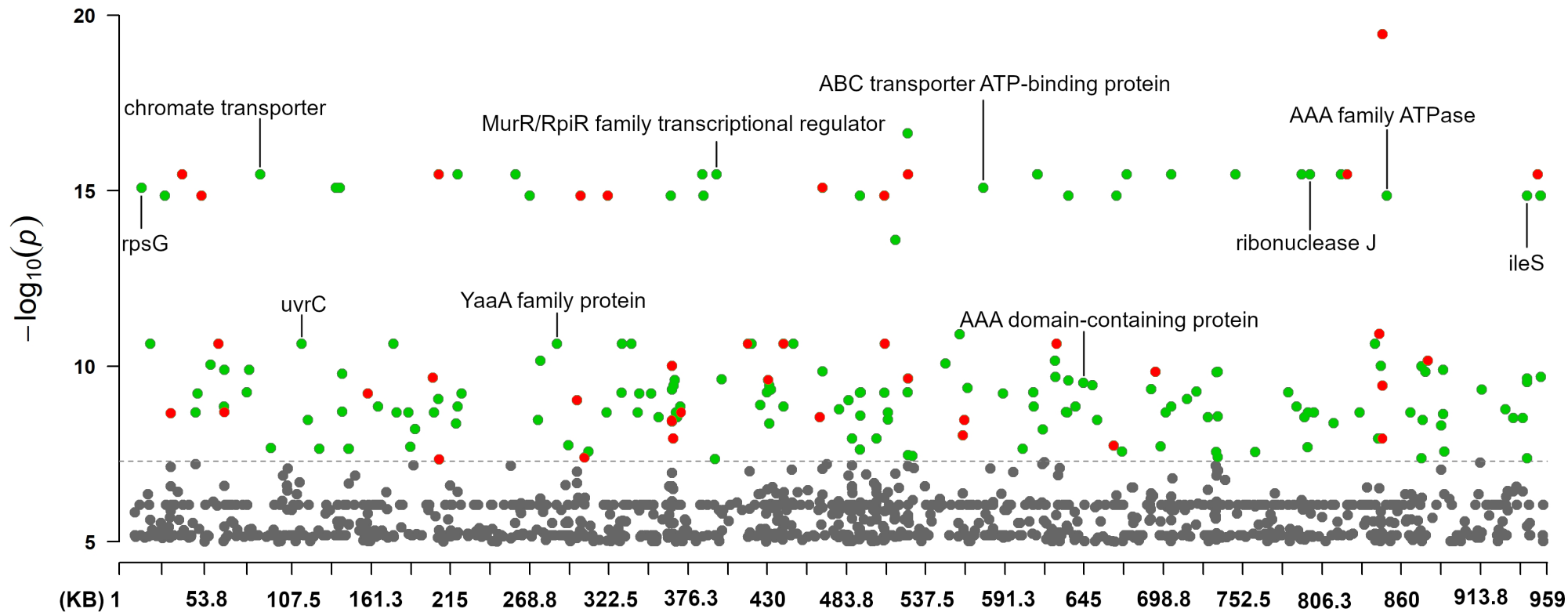
534

535 **Supplementary table S1. Background information of the *Mycoplasma anserisalpingtonis***
536 **strains.** The table contains the metadata of the strains the MIC values of the strains can be
537 found. Abbreviations: Na – data not available, bp – basepairs; GC% - guanine-cytosine
538 percentage of the genomes. The accession numbers in the table are the NCBI SRA accession
539 numbers.

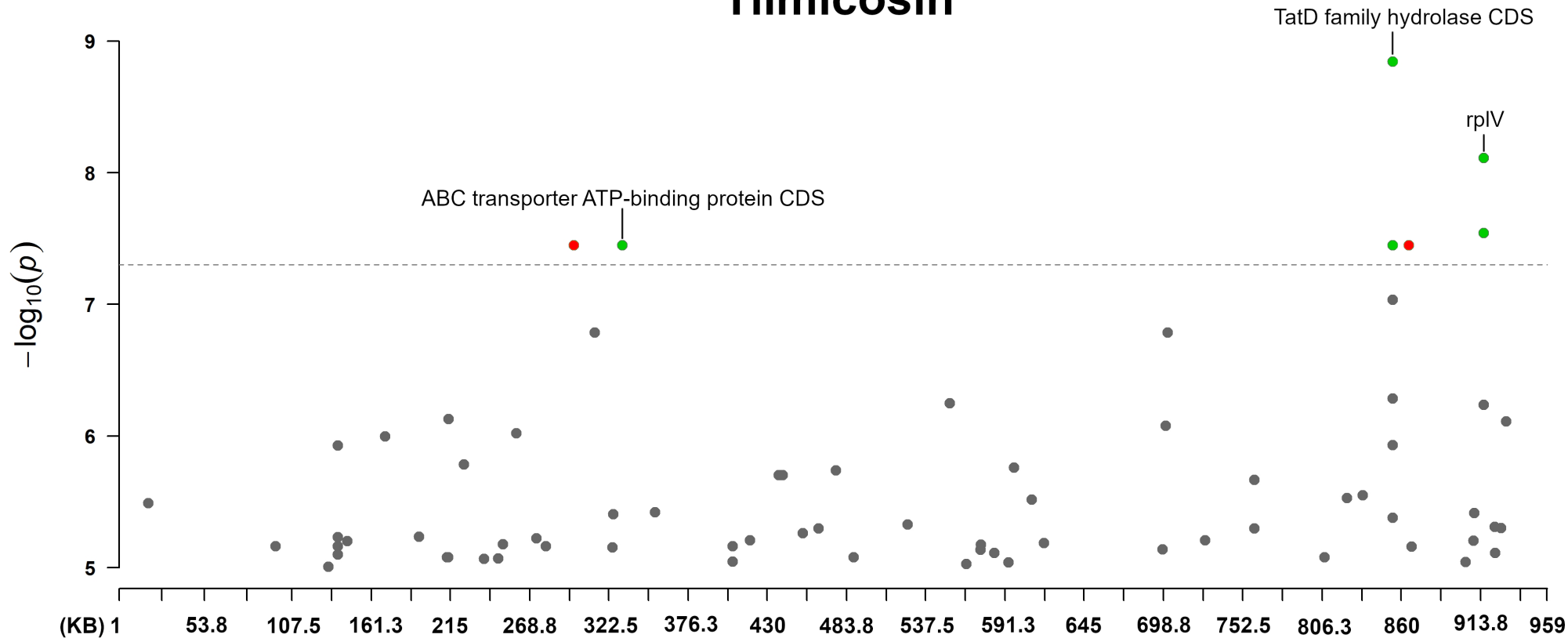
540 **Supplementary figure 1. Bar graphs of the MIC values of the antibiotic agents.** The bar graphs
541 of the antimicrobial agents. The X axis denotes the MIC values, while the Y axis denotes the
542 number of strains that were inhibited at that specific concentration.

543 **Supplementary data:** The draft genomes and the CDSs with significant k-mer hit have been
544 uploaded to the Figshare site (see Availability of data and materials)

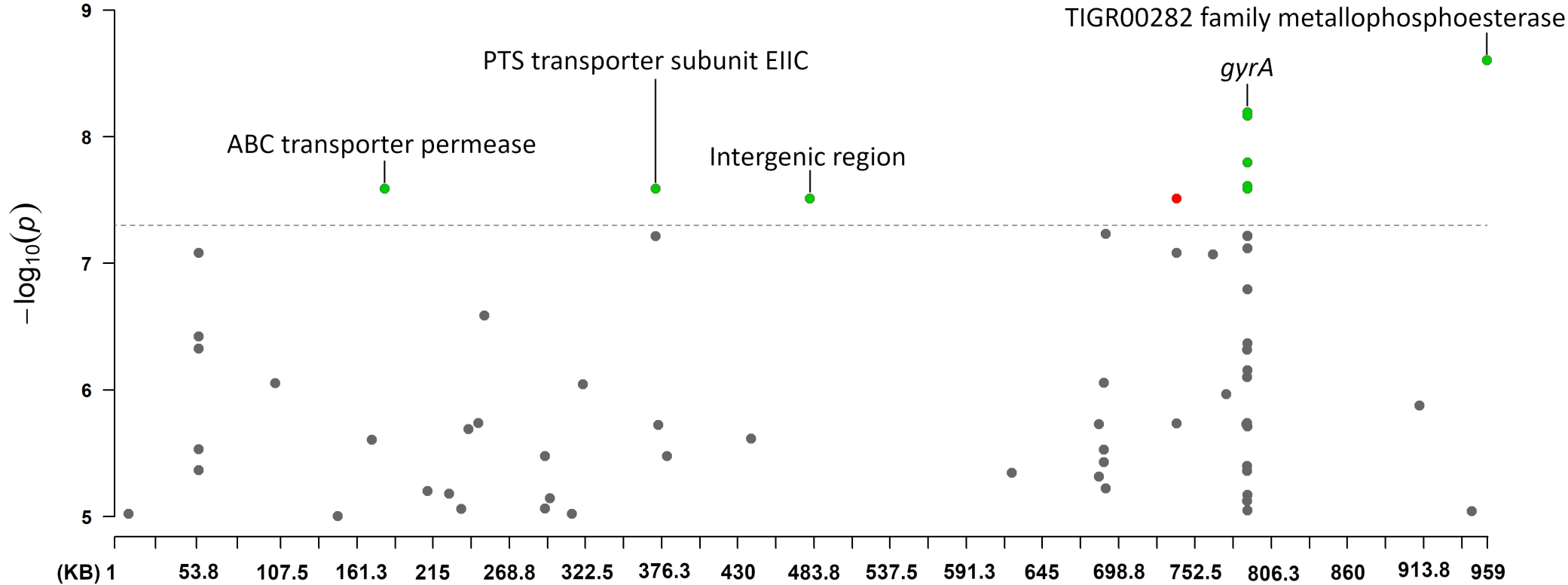
Tylvalosin



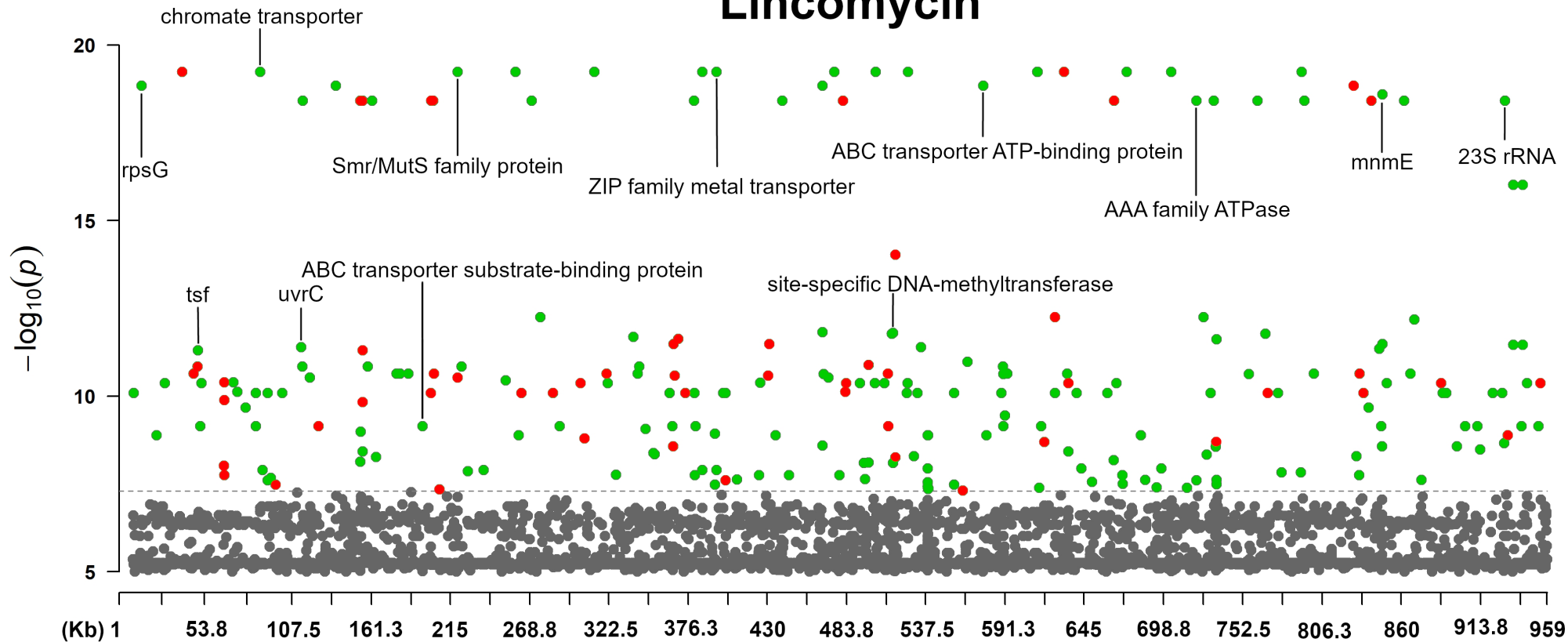
Tilmicosin



Enrofloxacin



Lincomycin



Spectinomycin

