

RESEARCH

Open Access



Genome-wide association study reveals serovar-associated genetic loci in *Riemerella anatipestifer*

Zhishuang Yang^{1,2,3}, Xueqin Yang¹, Mingshu Wang^{1,2,3,4}, Renyong Jia^{1,2,3,4}, Shun Chen^{1,2,3,4}, Mafeng Liu^{1,2,3,4}, Xinxin Zhao^{1,2,3,4}, Qiao Yang^{1,2,3,4}, Ying Wu^{1,2,3,4}, Shaqiu Zhang^{1,2,3,4}, Juan Huang^{1,2,3,4}, Xumin Ou^{1,2,3,4}, Sai Mao^{1,2,3,4}, Qun Gao^{1,2,3,4}, Di Sun^{1,2,3,4}, Bin Tian^{1,2,3,4}, Dekang Zhu^{1,2,3,4*} and Anchun Cheng^{1,2,3,4*}

Abstract

Background The disease caused by *Riemerella anatipestifer* (*R. anatipestifer*, RA) results in large economic losses to the global duck industry every year. Serovar-related genomic variation, such as the O-antigen and capsular polysaccharide (CPS) gene clusters, has been widely used for serotyping in many gram-negative bacteria. RA has been classified into at least 21 serovars based on slide agglutination, but the molecular basis of serotyping is unknown. In this study, we performed a pan-genome-wide association study (Pan-GWAS) to identify the genetic loci associated with RA serovars.

Results The results revealed a significant association between the putative CPS synthesis gene locus and the serological phenotype. Further characterization of the CPS gene clusters in 11 representative serovar strains indicated that they were highly diverse and serovar-specific. The CPS gene cluster contained the key genes *wzx* and *wzy*, which are involved in the Wzx/Wzy-dependent pathway of CPS synthesis. Similar CPS loci have been found in some other species within the family *Weeksellaceae*. We have also shown that deletion of the *wzy* gene in RA results in capsular defects and cross-agglutination.

Conclusions This study indicates that the CPS synthesis gene cluster of *R. anatipestifer* is a serotype-specific genetic locus. Importantly, our finding provides a new perspective for the systematic analysis of the genetic basis of the *R. anatipestifer* serovars and a potential target for establishing a complete molecular serotyping scheme.

Keywords *Riemerella anatipestifer*, Pan-genome wide association studies, Capsule, Serovar-associated

*Correspondence:

Dekang Zhu
zdk24@sicau.edu.cn

Anchun Cheng
chenganchun@vip.163.com

¹ Research Center of Avian Diseases, College of Veterinary Medicine, Sichuan Agricultural University, Chengdu, Sichuan, China

² Key Laboratory of Animal Disease and Human Health of Sichuan Province, Chengdu, Sichuan, China

³ International Joint Research Center for Animal Disease Prevention and Control of Sichuan Province, Chengdu, Sichuan, China

⁴ Engineering Research Center of Southwest Animal Disease Prevention and Control Technology, Ministry of Education of the People's Republic of China, Chengdu, Sichuan, China

Background

Riemerella anatipestifer (*R. anatipestifer*, RA) is an important veterinary pathogen belonging to the family *Weeksellaceae* [1]. RA poses a threat to domestic ducks, geese, and turkeys, causing acute or chronic septicemia [2]. Since 1982, when Bisgaard established the *R. anatipestifer* serotyping scheme [3], at least 21 serovars have been reported around the world [4]. And there was no cross-reactivity between different serovars [5]. There are several sets of reference strains that have been used for serovar designation by different groups [3, 6, 7]. Recently,



© The Author(s) 2024. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Omaleki et al. reconfirmed the serovar reference strains and identified 17 different serovars, which is the latest validated serotyping scheme [8]. Nonetheless, the serotyping of *R. anatipestifer* still relies heavily on reference strains and antisera, and no molecular serotyping methods have been proposed.

In most bacteria, the surface polysaccharide structures exhibit intraspecies diversity, which is usually associated with serological phenotype [9, 10]. The capsular polysaccharide (CPS) synthesis gene cluster was frequently utilized as a molecular serotyping target due to its genetic diversity [9, 11, 12]. Notably, due to the high correlation between the genetic signature of CPS gene cluster and the serovar phenotype, CPS related genes have been widely used as targets for molecular serotyping of many bacteria [12–14]. The most commonly targeted genes are *wzx* and *wzy*, which encode the oligosaccharide unit flippase (Wzx) and polymerase (Wzy), respectively. This gene pair plays a crucial role in CPS synthesis, which is directly linked to the specificity of CPS [11, 13]. Several studies have focused on the genes that are related to lipopolysaccharide (LPS) and CPS in *R. anatipestifer* [15–17], but the information they provide regarding serovar characteristics is limited. Previous studies have speculated that the serological characteristics of *R. anatipestifer* are associated with surface polysaccharides [18]. Our recent study indicates that the CPS of *R. anatipestifer* determines the serological specificity of serovar 2, rather than the LPS [19], while the genetic loci determining the serovars of *R. anatipestifer* remain unclear.

With the development of whole-genome sequencing technology and the accumulation of genomic data,

genome-wide association studies (GWAS) have become a powerful tool in bacterial research to reveal the genetic basis of important phenotypes. Recently, several methods have been introduced to assess the correlation between bacterial phenotypes and genotypes [20, 21]. Among these, the pan-genome-wide association study (Pan-GWAS), based on the typical gene presence/absence features of bacterial accessory genomes, has proven effective in studies of bacterial resistance and pathogenicity [22, 23]. There have been several pan-genomic studies of *R. anatipestifer* [24, 25]. However, to the best of our knowledge, no study has used the pan-GWAS approach to determine the association between genotypes and phenotypes in *R. anatipestifer*.

In the current study, we used pan-GWAS to identify the genetic loci associated with serovars. And we further analyzed the biological functions of the serologically associated genetic loci and preliminarily characterized our findings using gene knockout methods. These findings will provide a basis for further exploration of the molecular mechanism of *R. anatipestifer* serological phenotypes and provide direction for the establishment of molecular serotyping methods.

Materials and methods

Bacterial strain and whole-genome sequencing

The *R. anatipestifer* strains and the published genome data employed in this study are listed in Supplementary Table 1. One representative strain of each serovar was chosen for further presentation. Serovar representative strains and serotyping information were shown in Table 1. And all representative strains were obtained

Table 1 Serovar representative strains in this study

CCUG ID of representative strains	Strains name	Serovar	Bisgaard (1982) Serovar Reference Strains [3]	Bisgaard (1982) Serovar [3]	Rubbenstroth et al. (2013) Reference strains [7]	Rubbenstroth et al. (2013) Serovar [7]	Omaleki et al. (2021) Bisgaard Serovar Reference Strains [8]	Omaleki et al. (2021) Bisgaard Serovar [8]
CCUG18373	P-892	U1			DRL-24105	1		
CCUG25001	HPRS 2591	2	HPRS 2591	2	DRL-24046	2	HPRS 2591	2
CCUG25002	HPRS 2212	3	HPRS 2212	3	DRL-26338	3	HPRS 2212	3
CCUG25004	HPRS 2514	5	HPRS 2514	5	DRL-24123	5	HPRS 2514	5
CCUG25005	HPRS 2336	6	HPRS 2336	6	P-2123	6	HPRS 2336	6
CCUG25008	HPRS 2174	8	HPRS 2174	8	HPRS-2199	10	HPRS 2174	8
CCUG25010	HPRS 2528	9	HPRS 2528	9	DRL-27179	7	HPRS 2528	9
CCUG25011	HPRS 2564	10	HPRS 2564	10	HPRS-2565	4	HPRS 2564	10
CCUG25013	Singapore 8	U2			DRL-28020	11		
CCUG25054	HPRS 2560	11	HPRS 2560	11	DRL-26220	8	HPRS 2560	11
CCUG25055	8755/9	12	8755/9	12	8755/9	12	8755/9	12

CCUG Culture Collection University of Gothenburg, U1 Undefined type 1, U2 Undefined type 2

from the Culture Collection University of Gothenburg (CCUG). Other strains of *R. anatipestifer* used in this study were identified, characterized and archived by the Research Center of Avian Diseases, Sichuan Agricultural University (Chengdu, Sichuan, China).

The *R. anatipestifer* strains were grown in tryptic soy broth (TSB) and tryptic soy agar (TSA), at 37°C for 12h under microaerophilic conditions. *R. anatipestifer* genomic DNA was extracted using the TIANamp Genomic DNA Kit (TIANGEN BIOTECH, China). Whole-genome sequencing was performed using the Illumina HiSeq 2500 platform at the Beijing Genomics Institute (BGI, Shenzhen, China). Short-reads were filtered by Fastp (v0.19.4, default settings) [26] and draft genomes were assembled using SPAdes (v 3.11.0, default parameters with --careful flag) [27].

All *R. anatipestifer* strains were confirmed by RA-specific, 16S rRNA PCR assays (universal primer pairs 27F and 1410R), and genomic average nucleotide identity, as described previously [7, 28, 29]. and non-redundant isolates (>2000 SNPs/INDELS) were identified using custom scripts based on the NUCmer program (version 4.0.0beta2, https://github.com/youngDouble/MUMmerSNP2VCF_script) [30].

Agglutination test using the antisera

The serovars of *R. anatipestifer* involved in this study were determined by slide agglutination according to Bisgaard [3]. Standard serotyping antisera were obtained from RIPAC-LABOR GmbH (Potsdam, Germany), and those antisera have been extensively tested to indicate no cross-reactivity [6, 7]. As our representative strains are mainly from Bisgaard's scheme [3], we have labelled the serovar types according to the recent study by Omaleki et al. [8]. For non-serovar reference strains, we have labelled them with temporary serovar types. The serovar represented by CCUG 18373 is labeled as undefined type 1 (U1) and the serovar represented by CCUG 25013 is labeled as undefined type 2 (U2) (Table 1).

Genome wide association study of *R. anatipestifer* serovar phenotypes

To explore the association between *R. anatipestifer* serovars and genetic characteristics, a pan-genome-wide association study (Pan-GWAS) was performed. To ensure statistical power, GWAS was performed for the three most prevalent serovars in China. Specifically, the *R. anatipestifer* genome was annotated using Prokka (version 1.14.6, default parameters) [31], and the pan-genome containing 45 strains of *R. anatipestifer* was reconstructed with Roary (version 3.12.0, with identity threshold of protein=90) [32]. Furthermore, Scoary (v1.6.16) [20] was used to perform the Pan-GWAS with

the *gene_presence_absence* file generated by Roary (only serovars containing more than 10 strains were considered). Scoary's *P*-value and *Q*-value (adjusted *P*-value, adjust algorithm: Benjamini-Hochberg method) cut-offs were set to 0.05, the sensitivity cut-off was set to 95% and specificity to 85%. Next, we mapped the genes that were significantly associated with the serovar to the corresponding genome to obtain the distribution characteristics. Contig comparisons were generated with Easyfig (v2.2) [33].

Functional speculation of the gene cluster

To explore the function of serovar-related genetic loci, genome-wide biosynthetic gene clusters (BGCs) of *R. anatipestifer* was predicted with antiSMASH (version 4.2.0, parameter setting: --clusterblast --subclusterblast --knownclusterblast --smcogs --inclusive --borderpredict) [34]. BGCs analysis was performed again by DeepBGC [35], which uses deep learning strategies to mine biosynthetic gene clusters in the microbial genome. The results of the above two methods will be considered comprehensive.

Gene boundary determination of *R. anatipestifer* CPS gene cluster

Based on the results of biosynthetic gene cluster mining, we further determined the boundaries of the *R. anatipestifer* CPS gene cluster. Specifically, we retrieved 450 known CPS gene clusters from the NCBI Nucleotide database (<https://www.ncbi.nlm.nih.gov/nucleotide>) (Supplementary Table 2). We downloaded the protein sequence of these gene clusters, used CD-HIT (version 4.8.1, parameter setting: -c 1 -aS 0.95) [36] to remove redundancies and constructed the database. TblastN (version 2.6+) was used to map these proteins to the *R. anatipestifer* genome, and the resulting filtering thresholds were as follows: coverage $\geq 50\%$ (-qcov_hsp_perc 50), e-value $\leq 1e-5$ (-evalue 1e-5). Subsequently, the densely mapped regions in the genome are considered as candidates for the CPS gene cluster. Finally, combined with the prediction results of BGCs, the boundary of the gene cluster was determined by manual inspection.

Annotation of the CPS synthesis gene cluster

Protein-encoding genes were predicted using Prokka [31] and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [37] with default parameters. To assign functions to the predicted genes, the Conserved Domains Database (CDD) [38] was used to search for conserved domains with an E-value threshold of 0.01. Meanwhile characteristic gene annotation of genes was performed using Blastp (v2.6+) against Non-Redundant (NR, <https://ftp.ncbi.nlm.nih.gov/blast/db/>) database. The E-value and query

coverage were set at $1e-5$ and 50% respectively. Wzx and Wzy are key proteins in CPS synthesis and possess a typical multi-transmembrane structure [39]. Therefore, for the identification of Wzx and Wzy, TMHMM2.0 [40] was used to predict the transmembrane regions of proteins.

Inter- and intra-serovars comparison of CPS gene cluster

Gene cluster nucleotide sequence alignment was performed using MAFFT [41] in automatic mode, and then Mega X [42] with default parameters and 1000 bootstrap replicates were used to reconstruct the Neighbor-joining (NJ) [43] phylogenetic tree.

Blast (v2.6+) and Easyfig (v2.2) [33] were used for inter- and intra-serovar CPS gene cluster comparisons. Clustal Omega Web services (<https://www.ebi.ac.uk/Tools/msa/clustalo/>, default settings) was used to calculate the percentage identity of all Wzx and Wzy protein sequences.

Conservation analysis of CPS locus in family *Weeksellaceae*

To investigate the conservation of the CPS locus of *R. anatipestifer* in closely related species within the family *Weeksellaceae*, the multi-gene search method was implemented against representative genome database (<https://ftp.ncbi.nlm.nih.gov/genomes/refseq/>, data as of November 12, 2022) [44]. Specifically, Multigeneblast [45] was used to find homologues of *R. anatipestifer* CPS gene cluster from the representative genomes of *Weeksellaceae* species (based on published data). In addition, we used Easyfig to compare the collinearity of the best homologues.

Construction of *R. anatipestifer* wzy mutant strain CH-2 Δ wzy

To further characterize the relationship between the predicted CPS gene cluster and capsule synthesis, we performed a deletion mutation in the predicted key gene. Briefly, the *wzy* gene (*G148_RS04365*) was deleted by allelic exchange using the recombinant suicide vector pYA4278 (Supplementary Fig. 1a, Kong et al. [46]; donated by Professor Kong). Briefly, upstream (L) and downstream (R) fragments of the *R. anatipestifer* CH-2 *wzy* gene were amplified by PCR from the genome using *wzy*-Left F and *wzy*-Left R, and *wzy*-Right F and *wzy*-Right R primers, respectively. A 1145-bp Spec^R cassette was PCR-amplified from the pYES1 new plasmid using the Spc F and Spc R primers. The three fragments were then spliced together in vitro by overlap extension using the *wzy*-Left F and *wzy*-Right R primers, producing the LSR fragment. Adenosine nucleotides were added to both ends of the PCR product, which was then ligated to the AhdI-digested T-cloning suicide vector pYA4278 to generate pYA4278-LSR, which carries a deletion of the entire *wzy* gene. Subsequently, pYA4278-LSR was successively

transformed into *E. coli* X7213 λ pir [47]. *E. coli* X7213 λ pir (Donor) and *R. anatipestifer* CH-2 (Recipient) were mixed in a 10 mM MgSO₄ solution and incubated on TSB agar with diaminopimelic acid at 37 °C for 24 h. Spec^R transconjugants were further selected in media containing spectinomycin (40 μ g/ml). The detailed steps of this study refer to the methods of Luo et al. [28]. To confirm the *R. anatipestifer* mutant CH-2 Δ wzy, we performed PCR targeting the transconjugants (see Supplementary Fig. 1b for details). The above strains and plasmids are preserved at the Research Center of Avian Diseases, College of Veterinary Medicine, Sichuan Agricultural University, Chengdu, Sichuan, China. The primers used for construction of the above strains and plasmids are listed in Table 2.

Microscopic imaging of capsules by India ink staining and transmission electron microscopy

Capsule staining of *R. anatipestifer* was carried out using an improved Indian ink staining method as previously described [48]. Briefly, one drop of Indian ink and one drop of bacterial suspension were mixed on a glass slide, spread thinly, and air-dried. The slide was then counter-stained with 1% crystal violet for 1 minute, gently rinsed with distilled water, air-dried, and observed by optical microscope.

The wild-type (CH-2) and mutant (CH-2 Δ wzy) strains were washed twice with phosphate-buffered saline at 5000 r/min for 10 minutes after overnight culture. The precipitate was treated with 2.5% glutaraldehyde (pH 7.2) for 2 hours. After washing 20 times with ultrapure water, the cells were adsorbed onto copper grids and then stained with the phosphotungstic acid solution for 5 minutes. The capsule was observed by field-emission transmission electron microscope (TEM, FEI Tecnai G2 F20, 200 kV).

Results

The serovars of *R. anatipestifer*

In this study, the genome data of *R. anatipestifer* involved a total of eleven serovars, including Serovar U1 ($n=10$), Serovar 2 ($n=11$), Serovar 3 ($n=1$), Serovar 4 ($n=1$), Serovar 5 ($n=1$), Serovar 6 ($n=3$), Serovar 9 ($n=1$), Serovar 11 ($n=1$), Serovar 8 ($n=5$), Serovar U2 ($n=10$), and Serovar 12 ($n=1$), which were determined by slide agglutination or from references. All strains and their information are shown in Supplementary Table 1.

The gene cluster associated with serovar phenotype of *R. anatipestifer*

To screen for loci associated with serovars, the GWAS was performed with Scoary on the serovars containing more than 10 strains (serovars U1, 2, U2). According to

Table 2 Primers for identification of *wzy* deletion

Name of primer	Targeted gene;Description	Product Length(bp)	Sequence	Source or reference
wzy-Left	G148_RS04360; Amplification of the <i>wzy</i> upstream fragment	609	F AAGAACATTACCCATATCCTATCGTTTCGA CGGTA	This study
			R TTCTGTCCTGGCTGGTTTTACGAATATTG TAAGATA	This study
wzy-Right	G148_RS04370; Amplification of the <i>wzy</i> downstream fragment	602	F CCAAGGTAGTCGGCAAATAATTTTATGAAA AAAGTAC	This study
			R TACATGAGAAACCACAAAAGCCTCTTTGGG AATA	This study
wzy	RA-CH-2 (G148_RS04365); Amplification of the <i>wzy</i>	886	F TCCAATGGGTTTACTTTCTTGTAACTTTGTCT	This study
			R CGTAATGGTTGGTTGAGATTCATTGGAG	This study
LSR	G148_RS04360 + spec; Identification of transconjugants	1199	F AGGTAGATAGGGCAAGTATGGCTTTTTCG	This study
			R ACCGTAACCAGCAAATCAATATCACTGTG	This study
16S rRNA_RA	16S rRNA; Identification of species	960	F CTTCGGATACTTGAGAGCG	[28]
			R GCAGCACCTTGAAAATTGT	[28]
Spec	<i>spec</i> ; Amplification of the <i>spec</i>	1145	F TCTTACAATATTCGTAAAACCAGCCAGGA CAGAAAT	This study
			R ACTTTTTTCATAAAATTATTGCGGACTAC CTTGGTG	This study

the Pan-GWAS filtering threshold, we obtained a total of 27 target genes, and each serovar harbors 9 associated genes (Fig. 1a, Supplementary Table 3). Next, we mapped these genes to the corresponding genome and found that these genes were close to each other and formed a gene cluster. Interestingly, according to the BGCs results predicted by antiSMASH, the gene clusters mentioned above were labelled as polysaccharide biosynthetic gene clusters. Furthermore, the presence of *wza* and *wzc* gene indicates that this gene cluster is responsible for the biosynthesis of CPS. (Supplementary Table 4). Based on these results, we speculate that the serovar-specific gene cluster was CPS biosynthesis gene cluster of *R. anatipestifer*.

We further compared the distribution of the gene cluster among different serovars, and the results showed that the position of the gene cluster was relatively conserved in the genome of *R. anatipestifer* (Fig. 2). In short, the gene region has conserved fragments of 4 and 5 genes at the beginning and end, respectively (Fig. 2).

To determine the boundaries of the CPS gene cluster, we focused on those locations with a high density of genes associated with CPS synthesis. The results show a distinct boundary in the region of the putative CPS gene cluster (Fig. 1b, Supplementary Fig. 2). Therefore, we speculate that the CPS gene cluster of *R. anatipestifer* is located between the regulatory protein coding gene *recX* and the ribosomal protein S12 methyltransferase coding gene *rimO*, both of which are highly conserved in the genome of *R. anatipestifer* (Fig. 2).

Inter- and intra-serovars comparison of CPS gene cluster

Based on the positional conservation of the CPS gene cluster, we extracted the CPS gene cluster sequences from serovar representative strains (Supplementary Table 5). The length of the gene clusters from 22.76 kb (serovar 8) to 30.18 kb (serovar U1), GC content between 32.55% (serovar 10) and 34.05% (serovar 2), which was significantly different from the genomic GC content (upper quartile: 34.98%, lower quartile: 34.82%, mean: 34.95%; paired t-test: p -value < 0.0001). We annotated the CPS gene clusters of the serovar representative strains, the results are shown in Supplementary Table 4. These gene clusters contain an average of 23 CDSs (ranging from 19 to 27). It is worth noting that all serovar CPS gene clusters contain *wza*, *wzc*, *wzx*, *wzy*, *rfaA*, *rfaB*, and *rfaC* genes. The presence of the set of *wzx* and *wzy* genes indicates that CPS synthesis in *R. anatipestifer* may be a Wzx/Wzy-related processing pathways (Supplementary Table 4, Supplementary Fig. 3).

Furthermore, a NJ phylogenetic tree based on the complete sequence of the CPS gene cluster and a synteny analysis of the CPS gene cluster (DNA sequence identity cut-off: 69%) are shown in Fig. 2. As expected, strains of the same serogroup have more similar CPS gene cluster structures to each other and clearly cluster together in the same phylogenetic clade. Except for a gene insertion event in CCUG25001, the genetic structure of the CPS gene cluster of all serovar 2 strains were highly similar (Fig. 2). According to the annotation results of Prokka and PGAP, the predicted function of

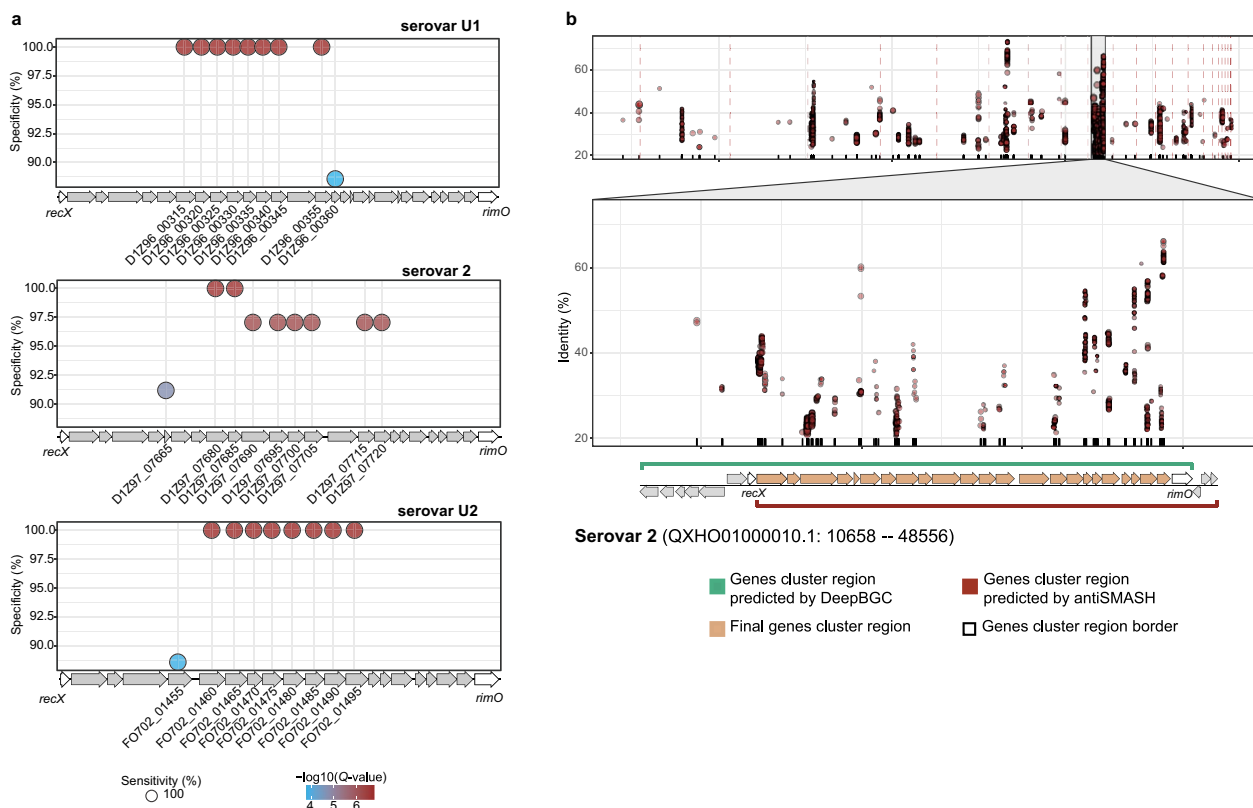


Fig. 1 Location of genes significantly associated with serovars. **a** And the specificity and the sensitivity of genes significantly associated with serovars U1, 2, and U2. The size of the shape indicates sensitivity; colour indicates negative log of adjusted *P*-value. **b** Gene cluster location and boundary determination. The dot plot represents the hits of genes related to CPS on the genome, and the size of the dot indicates the coverage length. Interval markers on gene clusters indicate the BGC regions predicted by DeepBGC and antiSMASH

the inserted gene is O-acetylase involved in peptidoglycan or LPS synthesis. Additionally, the cluster of serovar 5 and those of serovar 2 differed by only two genes (*wzx* and a glycosyltransferase coding gene). The gene clusters of serovar U1, 6, U2 strains have considerable similarity within the serogroup. Despite the relative diversity of the gene clusters of serovar 8 strains, their *wzx* and *wzy* are also identical. Comparative phylogenetic analysis revealed that the evolutionary trends of the core genome is inconsistent with CPS gene clusters (Fig. 2).

Next, we analysed the identity of the Wzx and Wzy sequences of all the strains by and constructed the NJ phylogenetic tree. Overall, Wzx and Wzy are serovar-specific, and much greater differences exist among the different serovars (Fig. 3a). Phylogenetic analysis of Wzx and Wzy, especially Wzx, clearly distinguishes the clades of different serovars. The minimum identity for the same serovars was 92.15% for Wzx and 95.14% for Wzy; the maximum identity for Wzx was 28.02% for different serovars and 26.50% for Wzy, except for serovars 5 with 2, which were 99.75% for Wzy (Fig. 3b).

Conserved loci in other *Weeksellaceae* species

The synteny analysis of homologous gene clusters in *Weeksellaceae* indicated that the locus of the CPS gene cluster was conserved among closely related species (Supplementary Fig. 4 and 5). Specifically, the upstream gene arrangement (*recx-gdr-wza-wzc*) of *R. anatipestifer* CPS gene cluster was highly conserved. *Chryseobacterium* and *R. anatipestifer* were the same (*recx* and *rimO*) at the beginning and end of the region.

As expected, this locus is also conserved in *Elizabethkingia sp.* and *Chryseobacterium sp.* (Supplementary Fig. 4b and c). Furthermore, many glycosyltransferases related to polysaccharide synthesis are distributed in this region in both genera. It is worth mentioning that *rfaA*, *rfaB*, and *rfaC* (*Elizabethkingia sp.*), LPS export system ATP-binding protein gene (*lptB*, *Elizabethkingia sp.*), ligase gene (*Chryseobacterium sp.*), and oligosaccharide flippase gene (*Chryseobacterium sp.*) were also present in the conserved region, and they are usually involved in the synthesis of CPS and LPS.

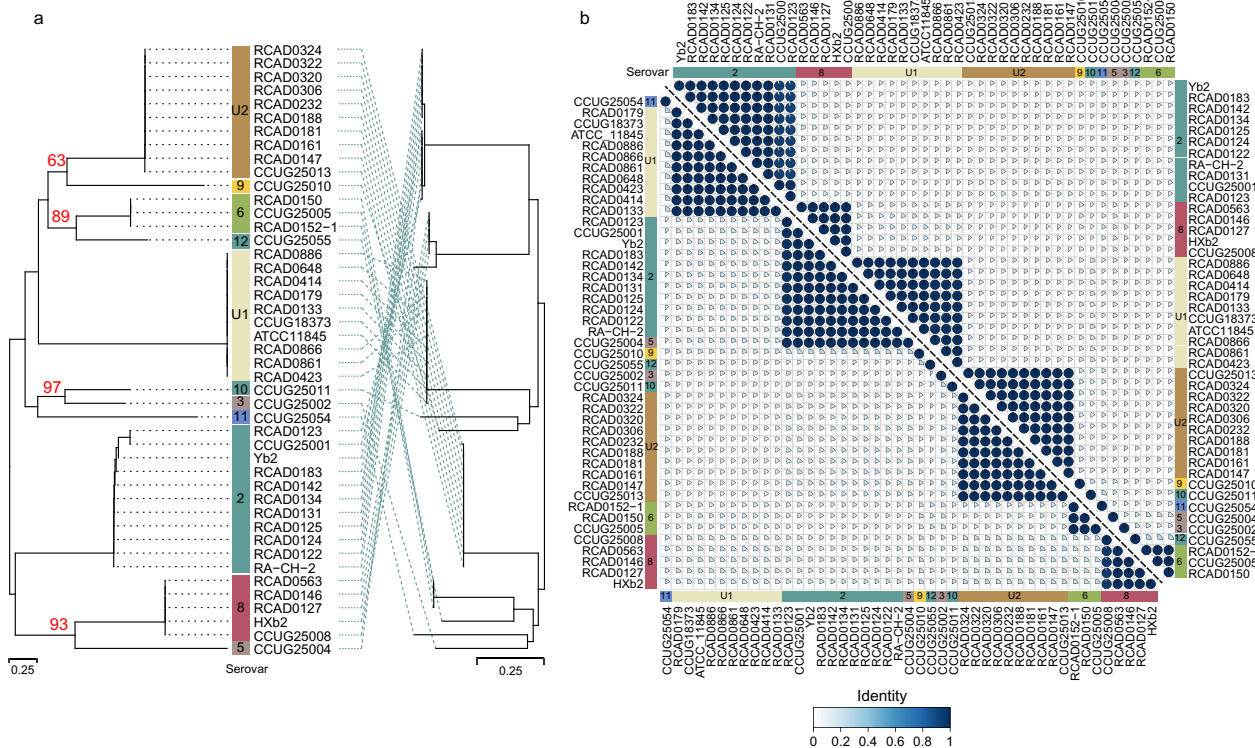


Fig. 3 **a** Phylogenetic tree constructed by the neighbor joining method based on the Wzx (left) and Wzy (right) protein sequences. **b** Identity of Wzx (upper triangle) and Wzy (lower triangle) sequences

and LSR fragment, but did not amplify the *wzy* fragment. All amplicons were confirmed by Sanger sequencing. After continuous culture for 30 generations, the genetic stability of the CH-2Δ*wzy* mutant was confirmed by the same PCR test (Supplementary Fig. 1c).

The results of India ink staining showed that there was a white ring-like capsule structure around the wild-type strain CH-2 (Fig. 4a), while there was none around the mutant strain CH-2Δ*wzy* (Fig. 4b). The capsule structure observed by transmission electron microscopy indicated that there was a layer of furry substance on the surface of the wild-type strain (Fig. 4c), and the capsule of the mutant strain (Fig. 4d) was thinner than the wild-type strain. Furthermore, the antisera slide agglutination test indicated that CH-2Δ*wzy* could agglutinate with the antisera from multiple serovars (Supplementary Fig. 6).

Discussion

Serotyping is an important method for characterizing *R. anatipestifer*, but corresponding molecular typing studies are still lacking. A recent study using time-of-flight mass spectrometry combined with machine learning algorithms has achieved recognition of serovars 1 and 2 [49]. Although the study was limited to a non-representative test set and did not indicate its corresponding molecular

basis, its results suggest the feasibility of molecular serotyping of *R. anatipestifer*.

Pan-GWAS have been applied in various bacteria to explore the genetic basis of various phenotypes [22, 23, 50]. In the present study, we used Pan-GWAS to identify the genetic loci significantly associated with three prevalent serovars of *R. anatipestifer*. Further functional analysis of the loci suggested that these loci are responsible for the synthesis of CPS. The result is in agreement with our recent finding that the CPS of *R. anatipestifer* determines the serological specificity of the serovar 2 strain [19]. This is also consistent with previous studies in other species, suggesting that each serovar corresponds to a specific CPS synthesis gene cluster [9, 51, 52].

Based on the results of the association study between serovar and genome, we predicted and analysed the CPS gene cluster of *R. anatipestifer*. We observed that the CPS gene cluster exhibits a genetic structure with highly conserved regions at both ends and a diversified middle region. The structure is similar to that of *Klebsiella pneumoniae*, *E. coli*, and *Acinetobacter baumannii*, and *Vibrio parahaemolyticus*, with typical genes (i.e., *wza* and *wzc*) located in the start region of the gene cluster [10, 12, 53, 54]. Despite some minor differences, the phylogenetic relationships between gene

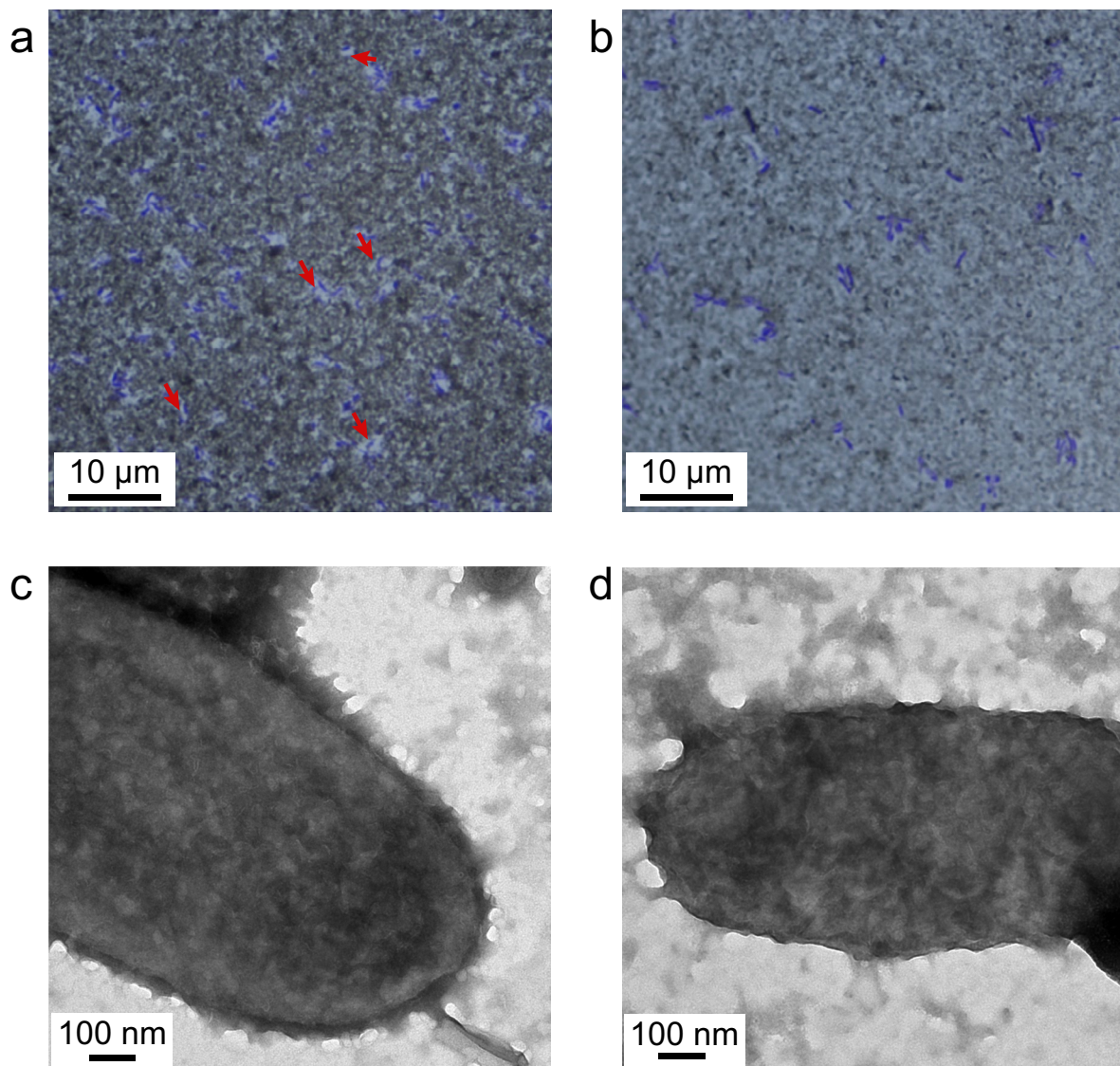


Fig. 4 Capsule staining (a, b) and transmission electron microscope (c, d). The capsule of the wild-type strain CH-2 (a) and mutant strain CH-2Δwzy (b) was stained and observed at 1000x magnification. The red arrows indicate the transparent capsule structures. The microstructure of wild-type strain CH-2 (c) and mutant strain CH-2Δwzy (d) at 130000x magnification

clusters of the same serovar are closer together. There are non-essential gene differences in the CPS gene cluster of serovar 9 strains of *Streptococcus suis*, but these differences did not cause phenotype change [55]. Considering the significant differences in isolation time and geographical location between recent *R. anatipestifer* isolates and the serovar reference strains, these differences in the CPS gene cluster appear to be explainable. The CPS gene clusters of serovars 2 and 5 differ by only three genes, of which *wzx* is one of the key genes determining serological specificity. This phenomenon is

also observed in other bacteria [10, 54, 56]. Notably, we observed phylogenetic inconsistency between the CPS gene clusters and the core genomes, which may be due to stronger selection pressure on capsule antigens [57]. The presence of *wzx* and *wzy* implies that the CPS of *R. anatipestifer* are processed via the Wzx/Wzy-dependent pathway [39]. *wzx* and *wzy* are widely used for capsule serotyping due to their excellent serovar specificity [12, 58], which has also been confirmed in our study. In *R. anatipestifer*, *wzx* can perfectly distinguish strains of different serovars, while *wzy* was slightly less effective as it cannot differentiate between serovar 2 and

5. Similar reports of two capsule serovars sharing the same *wzy* gene have been described in *Klebsiella spp* [56].

In this study, we also performed a conservative analysis of the CPS synthesis gene cluster of other *Weeksellaceae* species. It is noteworthy that a similar genetic locus is harbored in some species of *Chryseobacterium* and *Elizabethkingia*. To the best of our knowledge, there are no evidence-based reports of CPS synthetic gene clusters in *Chryseobacterium* and *Elizabethkingia*. Despite this limitation, we found several genes related to polysaccharide synthesis in these regions, such as *wbpA*, *wbpD*, *wbpE*, *lptB* and the ABC transporter ATP-binding protein gene [59, 60]. This is consistent with the previously putative capsular polysaccharide synthesis gene cluster of *Elizabethkingia* species [61]. Therefore, for some species of *Chryseobacterium* and *Elizabethkingia*, the above mentioned genomic region may also be the locus of the CPS synthesis gene cluster.

Furthermore, *wzy* gene (*G148_RS04365*) was deleted from *R. anatipestifer* CH-2. The absence of the capsule suggests that the *wzy* gene plays a crucial role in capsule synthesis, and previous study have shown that inactivation of *wza* in the CPS gene cluster also leads to the same phenotype [16]. Another study showed that knocking out the *AS87_04050* gene (coding Vi polysaccharide biosynthesis protein) from the CPS locus of *R. anatipestifer* Yb2 can alter serological characteristics, and the mutant strains exhibit a rough morphology [62], which is a typical feature of capsule loss [63]. And in the present study, the mutant strain CH-2 Δ *wzy* could agglutinate with antisera from multiple serovar. Similarly, deletion of the *M949_1603* gene (coding glycosyltransferase family 2 protein) in *R. anatipestifer* CH-1 results in cross-reactivity [15]. And we have confirmed that *M949_1603* gene is also located in the CPS gene cluster region. One possible explanation for the phenomenon of cross-reactivity is that the absence of capsule leads to the exposure of highly conserved epitopes. These results may indicate that the CPS of *R. anatipestifer* is the major antigenic component related to serological characteristics.

Conclusion

In this work, we revealed that association between the putative CPS gene cluster and the serovar types of *R. anatipestifer* through a genome-wide association studies. The CPS synthesis gene cluster of *R. anatipestifer* is serovar-specific. Moreover, the inactivation of the *wzy* gene results in defective capsule phenotype and cross-agglutination. This study provides new insights into the molecular basis of serotyping in *R. anatipestifer* and provides ideas for the development of molecular serotyping methods.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-09988-4>.

Additional file 1: Supplementary Table 1. The strains and the published genome data used in this study. **Supplementary Table 2.** Information of 450 known capsular polysaccharide synthesis gene cluster from the NCBI. **Supplementary Table 3.** The specificity and the sensitivity of genes significantly associated with serovar U1, 2, and U2. **Supplementary Table 4.** Functional prediction of the gene in capsular polysaccharide synthesis gene cluster of serovar representative strain. **Supplementary Table 5.** Location, length and GC content of capsular polysaccharide synthesis gene cluster of serovar representative strain.

Additional file 2: Supplementary Figure 1. a) Construction of *R. anatipestifer* *wzy* mutant strain CH-2 Δ *wzy*. b) Identification of *R. anatipestifer* CH-2 Δ *wzy*. Lane M: DL2000 DNA Marker; Lanes 1-3: 16S rRNA F and 16S rRNA R, which amplify a 960 bp fragment from *R. anatipestifer* 16S rRNA. Order: Wild-type(CH-2), mutant(CH-2 Δ *wzy*), and negative control (distilled water); Lanes 4-6: Spec F and Spec R, which amplify a 1180 bp fragment from the SpecR cassette. Order: Positive control (pYES1 new), mutant (CH-2 Δ *wzy*), and negative control (distilled water); Lanes 7-9: *wzy* F and *wzy* R, which amplify an 886 bp fragment from the *wzy* gene. Order: Wild-type(CH-2), mutant (CH-2 Δ *wzy*), and negative control (distilled water); Lanes 10-11: LSR F and LSR R, which amplify a 1199 bp fragment from the SpecR cassette, indicating that it was inserted in the correct position in the *R. anatipestifer* CH-2 genome. Order: Mutant (CH-2 Δ *wzy*), Negative control (distilled water). c) Identification of *R. anatipestifer* CH-2 Δ *wzy* after continuous culture for 30 generations. Lane M1: DL15000 DNA Marker; Lane M2: DL2000 DNA Marker. The rest of the lanes are identical to (b). **Supplementary Figure 2.** Gene cluster location and boundary determination of serovar U1 and serovar U2. The dot plot represents the hits of genes related to CPS on the genome, and the size of the dot indicates the coverage length. Interval markers on gene clusters indicate the BGC regions predicted by DeepBGC and antiSMASH. **Supplementary Figure 3.** The prediction of transmembrane helices in amino acid sequences encoded by *wzx* and *wzy*. **Supplementary Figure 4.** Conserved loci in other *Weeksellaceae* species. a) The genetic locus of the CPS biosynthesis gene cluster in *R. anatipestifer* is conserved among the closest species. b) Conserved structure in multiple *Elizabethkingia* species. c) Conserved structure in multiple *Chryseobacterium* species. d) Comparison of the CPS gene cluster between *R. anatipestifer* (CH-2) and *R. columbina* (DSM 16469). **Supplementary Figure 5.** 16S rDNA NJ phylogenetic tree of closely related species (representative genome from NCBI) for *R. anatipestifer*. 16S rDNA nucleotide sequence alignment was performed using MAFFT and tree was reconstructed by MEGA X with default parameters and 1000 bootstrap replicates. **Supplementary Figure 6.** Agglutination test of *R. anatipestifer* mutant CH-2 Δ *wzy*. The Mutants are capable of reacting to multiple antisera. And the details of the reaction differed from the wild type.

Acknowledgements

Not applicable.

Authors' contributions

ZY designed the work, performed data analysis, and wrote the manuscript. XY assisted with molecular experiments. MW, RJ, SC, and ML managed and supervised the project. XZ, QY, YW, SZ, JH, XO, SM, QG, DS, and BT reviewed and revised the manuscript. DZ and AC conceptualized and designed the work, provided funding support, and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Sichuan Science and Technology Program (2020YJ0330); Sichuan Veterinary Medicine and Drug Innovation Group of China Agricultural Research System (SCCXTD-2021-18); Earmarked Fund for China Agriculture Research System (CARS-42-17).

Availability of data and materials

All genomic data generated or analysed in this study are archived in the NCBI Genome Database (<https://www.ncbi.nlm.nih.gov/assembly/>) and accession numbers can be found in Supplementary Table 1.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 26 April 2023 Accepted: 8 January 2024

Published online: 13 January 2024

References

- García-López M, Meier-Kolthoff JP, Tindall BJ, Gronow S, Woyke T, Kyrpides NC, et al. Analysis of 1,000 type-strain genomes improves taxonomic classification of Bacteroidetes. *Front Microbiol.* 2019;10:2083.
- Boulianne M, Blackall PJ, Hofacre CL, Ruiz JA, Sandhu TS, Hafez HM, et al. *Pasteurellosis* and other respiratory bacterial infections. In: *Diseases of poultry*. 14th ed. Wiley; 2020. p. 831–89.
- Bisgaard M. Antigenic studies on *pasteurella anatipestifer*, species incertae sedis, using slide and tube agglutination. *Avian Pathol.* 1982;11(3):341–50.
- Pathanasophon P, Phuektes P, Tanticharoenyos T, Narongsak W, Sawada T. A potential new serotype of *Riemerella anatipestifer* isolated from ducks in Thailand. *Avian Pathol.* 2002;31(3):267–70.
- Liao H, Cheng X, Zhu D, Wang M, Jia R, Chen S, et al. TonB energy transduction Systems of *Riemerella anatipestifer* are required for Iron and hemin utilization. *PLoS One.* 2015;10(5):e0127506.
- Sandhu TS, Leister ML. Serotypes of '*Pasteurella*' *anatipestifer* isolates from poultry in different countries. *Avian Pathol.* 1991;20(2):233–9.
- Rubbenstroth D, Ryll M, Knobloch JK-M, Köhler B, Rautenschlein S: evaluation of different diagnostic tools for the detection and identification of *Riemerella anatipestifer*. *Avian Pathol.* 2013;42(1):17–26.
- Omaleki L, Blackall PJ, Bisgaard M, Turni C. Molecular and serological characterization of *Riemerella* isolates associated with poultry in Australia. *Avian Pathol.* 2021;50(1):31–40.
- Bossé JT, Li Y, Fernandez Crespo R, Lacouture S, Gottschalk M, Sárközi R, et al. Comparative sequence analysis of the capsular polysaccharide loci of *Actinobacillus pleuropneumoniae* serovars 1–18, and development of two multiplex PCRs for comprehensive capsule typing. *Vet Microbiol.* 2018;220:83–9.
- Bian S, Zeng W, Li Q, Li Y, Wong N-K, Jiang M, et al. Genetic structure, function, and evolution of capsule biosynthesis loci in *Vibrio parahaemolyticus*. *Front Microbiol.* 2020;11:546150.
- Weinberger DM, Trzciński K, Lu Y-J, Bogaert D, Brandes A, Galagan J, et al. *Pneumococcal* capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog.* 2009;5(6):e1000476.
- Pan Y-J, Fang H-C, Yang H-C, Lin T-L, Hsieh P-F, Tsai F-C, et al. Capsular polysaccharide synthesis regions in *Klebsiella pneumoniae* serotype K57 and a new capsular serotype. *J Clin Microbiol.* 2008;46(7):2231–40.
- Li S-C, Huang J-F, Hung Y-T, Wu H-H, Wang J-P, Lin J-H, et al. In silico capsule locus typing for serovar prediction of *Actinobacillus pleuropneumoniae*. *Microbial Genomics.* 2022;8(4):000780.
- Klima CL, Zaheer R, Briggs RE, McAllister TA. A multiplex PCR assay for molecular capsular serotyping of *Mannheimia haemolytica* serotypes 1, 2, and 6. *J Microbiol Methods.* 2017;139:155–60.
- Zou J, Wang X, Ding C, Tian M, Han X, Wang S, et al. Characterization and cross-protection evaluation of M949_1603 gene deletion *Riemerella anatipestifer* mutant RA-M1. *Appl Microbiol Biotechnol.* 2015;99(23):10107–16.
- Yi H, Yuan B, Liu J, Zhu D, Wu Y, Wang M, et al. Identification of a wza -like gene involved in capsule biosynthesis, pathogenicity and biofilm formation in *Riemerella anatipestifer*. *Microb Pathog.* 2017;107:442–50.
- Dou Y, Yu G, Wang X, Wang S, Li T, Tian M, et al. The *Riemerella anatipestifer* M949_RS01035 gene is involved in bacterial lipopolysaccharide biosynthesis. *Vet Res.* 2018;49(1):93.
- Brogden KA, Rhoades KR, Rimler RB. Serologic types and physiologic characteristics of 46 avian *Pasteurella anatipestifer* cultures. *Avian Dis.* 1982;7.
- Liu Y, Luo S, Yang Z, Wang M, Jia R, Chen S, et al. Capsular polysaccharide determines the serotyping of *Riemerella anatipestifer*. *Microbiol Spectrum.* 2023:e01804–23.
- Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* 2016;17(1):238.
- Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. Pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics.* 2018;3.
- Yuan J, Li Y-Y, Xu Y, Sun B-J, Shao J, Zhang D, et al. Molecular signatures related to the virulence of *Bacillus cereus* Sensu Lato, a leading cause of devastating Endophthalmitis. *Msystems.* 2019;4(6):12.
- Ma KC, Mortimer TD, Duckett MA, Hicks AL, Wheeler NE, Sánchez-Busó L, et al. Increased power from conditional bacterial genome-wide association identifies macrolide resistance mutations in *Neisseria gonorrhoeae*. *Nat Commun.* 2020;11(1):5374.
- Zhu D, Yang Z, Xu J, Wang M, Jia R, Chen S, et al. Pan-genome analysis of *Riemerella anatipestifer* reveals its genomic diversity and acquired antibiotic resistance associated with genomic islands. *Function Integr Genomics.* 2020;20(3):307–20.
- Liu J, Zeng Q, Wang M, Cheng A, Liu M, Zhu D, et al. Comparative genome-scale modelling of the pathogenic Flavobacteriaceae species *Riemerella anatipestifer* in China. *Environ Microbiol.* 2019;21(8):2836–51.
- Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–90.
- 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(5):455–77.
- Luo H, Liu M, Wang L, Zhou W, Wang M, Cheng A, et al. Identification of ribosomal RNA methyltransferase gene *ermF* in *Riemerella anatipestifer*. *Avian Pathol.* 2015;44(3):162–8.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun.* 2018;9(1):5114.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):1–9.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30(14):2068–9.
- Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015;31(22):3691–3.
- Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics.* 2011;27(7):1009–10.
- Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. Weber T: antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 2019;47(W1):W81–7.
- Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* 2019;47(18):e110–0.
- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22(13):1658–9.
- Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, et al. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* 2021;49(D1):D1020–8.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015;43(D1):D222–6.
- Whitfield C, Wear SS, Sande C. Assembly of bacterial capsular polysaccharides and exopolysaccharides. *Annu Rev Microbiol.* 2020;74(1):521–43.

40. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Mol Biol.* 2001;305(3):567–80.
41. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–66.
42. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
43. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4(4):406–25.
44. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* 2014;42(D1):D553–9.
45. Medema MH, Takano E, Breitling R, Bankevich A, Nurk S, Antipov D, et al. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol.* 2013;30(5):1218–23.
46. Leclercq R, Courvalin P. Bacterial resistance to macrolide, lincosamide, and streptogramin antibiotics by target modification. *Antimicrob Agents Chemother.* 1991;35(7):1267–72.
47. Edwards RA, Keller LH, Schifferli DM. Improved allelic exchange vectors and their use to analyze 987P fimbria gene expression. *Gene.* 1998;207(2):149–57.
48. Breakwell DP, Moyes RB, Reynolds J. Differential staining of Bacteria: capsule stain. *Curr Protoc Microbiol.* 2009;15(1)
49. Wang Z, Zheng X, Chen J, Xu Z, Dong Y, Xu G, et al. Machine learning combined with MALDI-TOF MS has the potential ability to identify serotypes of the avian pathogen *Riemerella anatipestifer*. *J Appl Microbiol.* 2022;124(1):e00075.
50. Mageiros L, Méric G, Bayliss SC, Pensar J, Pascoe B, Mourkas E, et al. Genome evolution and the emergence of pathogenicity in avian *Escherichia coli*. *Nat Commun.* 2021;12(1):765.
51. LaClaire LL, Tondella MLC, Beall DS, Noble CA, Raghunathan PL, Rosenstein NE, et al. Identification of *Haemophilus influenzae* serotypes by standard slide agglutination serotyping and PCR-based capsule typing. *J Clin Microbiol.* 2003;41(1):393–6.
52. Harrison OB, Claus H, Jiang Y, Bennett JS, Bratcher HB, Jolley KA, et al. Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerg Infect Dis.* 2013;19(4):566–73.
53. Whitfield C. Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem.* 2006;75(1):39–68.
54. Wyres KL, Cahill SM, Holt KE, Hall RM, Kenyon JJ. Identification of *Acinetobacter baumannii* loci for capsular polysaccharide (KL) and lipooligosaccharide outer core (OCL) synthesis in genome assemblies using curated reference databases compatible with Kaptive. *Microbial Genomics.* 2020;6(3):e000339.
55. Okura M, Takamatsu D, Maruyama F, Nozawa T, Nakagawa I, Osaki M, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters from all serotypes of *Streptococcus suis*: potential mechanisms for generation of capsular variation. *Appl Environ Microbiol.* 2013;79(8):2796–806.
56. Pan Y-J, Lin T-L, Chen C-T, Chen Y-Y, Hsieh P-F, Hsu C-R, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of *Klebsiella spp.* *Sci Rep.* 2015;5(1):15573.
57. Mostowy RJ, Holt KE. Diversity-generating machines: genetics of bacterial sugar-coating. *Trends Microbiol.* 2018;26(12):1008–21.
58. Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *J Clin Microbiol.* 2006;44(1):124–31.
59. Luo Q, Yang X, Yu S, Shi H, Wang K, Xiao L, et al. Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. *Nat Struct Mol Biol.* 2017;24(5):469–74.
60. Shoji M, Sato K, Yukitake H, Naito M, Nakayama K. Involvement of the Wbp pathway in the biosynthesis of *Porphyromonas gingivalis* lipopolysaccharide with anionic polysaccharide. *Sci Rep.* 2014;4:5056.
61. Breurec S, Criscuolo A, Diancourt L, Rendueles O, Vandebogaert M, Passet V, et al. Genomic epidemiology and global diversity of the emerging bacterial pathogen *Elizabethkingia anophelis*. *Sci Rep.* 2016;6(1):30379.
62. Wang X, Ding C, Wang S, Han X, Hou W, Yue J, et al. The AS87_04050 gene is involved in bacterial lipopolysaccharide biosynthesis and pathogenicity of *Riemerella anatipestifer*. *PLoS One.* 2014;9(10):e109962.
63. MacLeod CM, Krauss MR. Stepwise intratype transformation of *pneumococcus* from R to S by way of a variant intermediate in capsular polysaccharide production. *J Exp Med.* 1947;86(6):439–52.

Publisher's Note

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