

In silico capsule locus typing for serovar prediction of *Actinobacillus pleuropneumoniae*

Siou-Cen Li^{1,2}, Jing-Fang Huang², Yu-Ting Hung², Hsiu-Hui Wu², Jyh-Perng Wang², Jiunn-Horng Lin², Zeng-Weng Chen^{2,*} and Shih-Ling Hsuan^{1,*}

Abstract

Actinobacillus pleuropneumoniae is a causative agent of pleuropneumonia in pigs of all ages. *A. pleuropneumoniae* is divided into 19 serovars based on capsular polysaccharides (CPSs) and lipopolysaccharides. The serovars of isolates are commonly determined by serological tests and multiplex PCR. This study aimed to develop a genomic approach for *in silico* *A. pleuropneumoniae* typing by screening for the presence of the species-specific *apxIV* gene in whole-genome sequencing (WGS) reads and identifying capsule locus (KL) types in genome assemblies. A database of the *A. pleuropneumoniae* KL, including CPS synthesis and CPS export genes, was established and optimized for Kaptive. To test the developed genomic approach, WGS reads of 189 *A. pleuropneumoniae* isolates and those of 66 samples from 14 other bacterial species were analysed. ARIBA analysis showed that *apxIV* was detected in all 189 *A. pleuropneumoniae* samples. These *apxIV*-positive WGS reads were *de novo* assembled into genome assemblies and assessed. A total of 105 *A. pleuropneumoniae* genome assemblies that passed the quality assessment were analysed by Kaptive analysis against the *A. pleuropneumoniae* KL database. The results showed that 97 assemblies were classified and predicted as 13 serovars, which matched the serovar information obtained from the literature. The six genome assemblies from previously nontypable isolates were typed and predicted as serovars 17 and 18. Notably, one of the two "*Actinobacillus porcitonisillarum*" samples was *apxIV* positive, and its genome assembly was typed as KL03 with high identity and predicted as *A. pleuropneumoniae* serovar 3. Collectively, a genomic approach was established and could accurately determine the KL type of *A. pleuropneumoniae* isolates using WGS reads. This approach can be used with high-quality genome assemblies for predicting *A. pleuropneumoniae* serovars and for retrospective analysis.

DATA SUMMARY

Supplementary data S1 and S2 and Tables S1 and S2 can be found on FigShare (<https://doi.org/10.6084/m9.figshare.19492493>) and with the online version of this article. The capsule locus (KL) sequence of *Actinobacillus pleuropneumoniae* serovar 6 strain Femo produced as part of this work is available from the National Center for Biotechnology Information GenBank with accession number MZ450073, and is shown in Supplementary data S2 as the KL06 reference sequence.

INTRODUCTION

Actinobacillus pleuropneumoniae, a Gram-negative bacterium, is a causative agent of porcine pleuropneumonia worldwide [1, 2]. *A. pleuropneumoniae* is currently classified into 19 serovars based on capsular polysaccharide (CPS, K antigen) and lipopolysaccharide [3, 4]. The main serovar determinant is CPS, production of which requires a functional capsule locus (KL) harbouring CPS synthesis (*cps*) genes and CPS export (*cpx*) genes [3, 5, 6].

Received 01 August 2021; Accepted 16 January 2022; Published 11 April 2022

Author affiliations: ¹Graduate Institute of Veterinary Pathobiology, College of Veterinary Medicine, National Chung Hsing University, Taichung, Taiwan, ROC; ²Animal Technology Research Center, Agricultural Technology Research Institute, Miaoli, Taiwan, ROC.

***Correspondence:** Zeng-Weng Chen, zwc@mail.atri.org.tw; Shih-Ling Hsuan, hsuan@nchu.edu.tw

Keywords: *Actinobacillus pleuropneumoniae*; capsule locus typing; genomic analysis.

Abbreviations: CPS, capsular polysaccharide; KL, capsule locus; NCBI, National Center for Biotechnology Information; ONT, Oxford Nanopore Technologies; rDNA, ribosomal DNA; SRA, Sequence Read Archive; WGS, whole-genome sequencing.

The GenBank/EMBL/DBJ accession number for the complete K locus sequence of *A. pleuropneumoniae* serovar 6 strain Femo (ATCC 33590) is MZ450073.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary data sets and two supplementary tables are available with the online version of this article and on FigShare.

000780 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Impact Statement

Actinobacillus pleuropneumoniae is an important respiratory pathogen of pigs. The prevalence of *A. pleuropneumoniae* serovars varies among countries and areas. Serotyping of *A. pleuropneumoniae* isolates provides information on the regional prevalence, which is the basis for disease prevention and control. In recent years, the use of comparative genomic analysis of capsular polysaccharide synthesis genes has helped scientists identify potential new *A. pleuropneumoniae* serovars, and molecular typing methods based on PCR have been improved accordingly. Nevertheless, some clinical *A. pleuropneumoniae* isolates remained nontypable, indicating that the diversity of the *A. pleuropneumoniae* capsule locus (KL) needed further investigation. This study developed a genomic approach to predict serovars through *in silico* KL typing using whole-genome sequencing (WGS) reads and bioinformatics tools from the public domain. This approach provides the compatibility of retrospective analysis and flexibility in an expandable database for new *A. pleuropneumoniae* serovars. It is important for increasing awareness toward emerging serovars and can be combined with other WGS-based analysis methods for genomic characterization of isolates.

Whole-genome sequencing (WGS) is emerging as a new technology for *A. pleuropneumoniae* studies, including for the prediction of antimicrobial resistance, identification of new serovars, and investigation of potential targets of drugs and vaccines [4, 7–10]. Through WGS and sequence analysis of *cps* genes, previously nontypable *A. pleuropneumoniae* isolates were determined to be new serovars 16–19 [4, 7, 8]. Molecular typing methods using multiplex PCR were modified accordingly to detect serovar-specific *cps* genes and the *A. pleuropneumoniae*-specific *apxIV* gene [4, 11]; however, serovars 9 and 11 remain undistinguishable. Comparative analysis of KLs between *A. pleuropneumoniae* serovars revealed differences in the coding sequences of proteins participating in CPS biosynthesis [4, 7, 8, 11], implying that *in silico* WGS data analysis of the KL may provide a higher resolution for typing. To the best of our knowledge, using whole-genome sequences to predict serovars has yet to be established in *A. pleuropneumoniae*.

Klebsiella is an antibiotic-resistant priority pathogen listed by the World Health Organization. Molecular typing of *Klebsiella* has relied on a multiplex PCR method targeting genes involved in the biosynthesis of surface polysaccharides over the past decade [12, 13]. Recently, WGS-based genotyping has been introduced to characterize the molecular types of *Klebsiella*. The genomic typing tool Kaptive, with both command-line and web interfaces, was developed for *Klebsiella* species, especially *Klebsiella pneumoniae*, to identify gene loci associated with biosynthesis of surface polysaccharides in a genome assembly [14, 15]. This WGS-based typing tool was integrated into genomic analysis workflows for pathogen characterization and epidemiological investigation [16, 17]. Kaptive is now used for genotyping *Klebsiella variicola* and *Acinetobacter baumannii* [16, 18], and whether Kaptive can be applied to *A. pleuropneumoniae* typing is of interest.

This study aimed to develop a genomic approach for the prediction of *A. pleuropneumoniae* serovars. Because genetic variation in the KL leads to diversity of the main serovar determinant, CPS, we tested whether the serovar of *A. pleuropneumoniae* isolates can be distinguished and predicted by *in silico* typing of KL sequences in genomes.

METHODS

In silico analysis workflow

The workflow of *in silico* typing of *A. pleuropneumoniae* is outlined in Fig. 1. Briefly, an *apxIV* dataset and a KL database were established. WGS short-read data without serovar information in the Sequence Read Archive (SRA) metadata of *A. pleuropneumoniae* and WGS short-read data of non-*A. pleuropneumoniae* bacteria were obtained from the National Center for Biotechnology Information (NCBI). Antimicrobial Resistance Identification By Assembly (ARIBA) and KL typing and variant evaluation (Kaptive) were used as tools to screen *A. pleuropneumoniae*-specific *apxIV* in WGS reads and to type the KL in the genome assembly, respectively. A KL type was assigned through Kaptive analysis, and the serovar of the *A. pleuropneumoniae* genome was predicted accordingly.

Establishment of an *apxIV* dataset

To establish an *apxIV* dataset compatible with ARIBA [19], a coding sequence of *ApXIV* (GenBank accession number AAD01698.1) with 1,805 amino acids from *A. pleuropneumoniae* serovar 1, strain 4074, was used as a reference sequence [20]. An associated metadata file describing information of the reference sequence was prepared to identify the presence or absence of the *apxIV* gene. The *apxIV* sequence and metadata are shown in Supplementary data S1.

Establishment of an *A. pleuropneumoniae* KL database

To establish an *A. pleuropneumoniae* KL reference database compatible with Kaptive [15], reference genomes of *A. pleuropneumoniae* serovars 1 to 5 and 7 to 12 and publicly available KL sequences of serovars 13 to 19 were collected and downloaded from the NCBI database (Table 1). The region of the full-length KL was identified by analysing the locations of the *modF* and *ydeN* genes

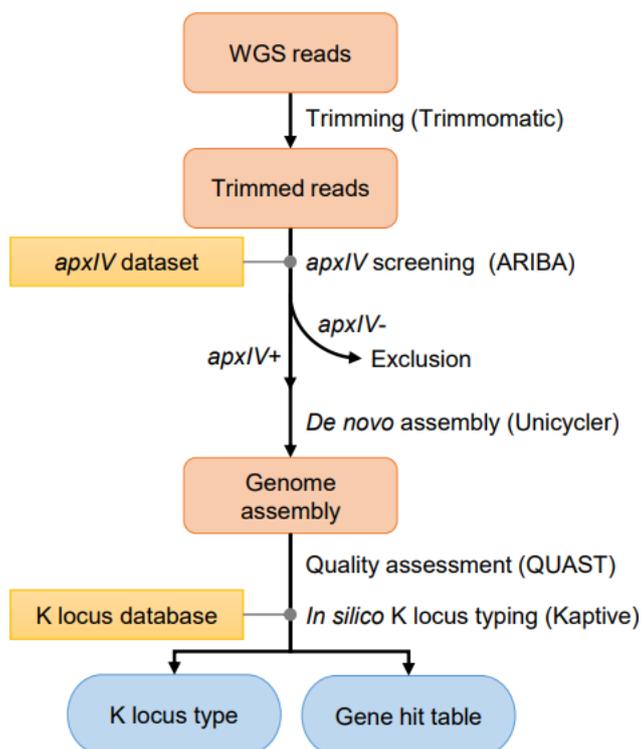


Fig. 1. Outline of the *in silico* KL typing of *A. pleuropneumoniae* using WGS reads. WGS short-read data were trimmed using Trimmomatic. The trimmed reads were analysed by ARIBA against the *apxIV* dataset to screen for the presence of the *A. pleuropneumoniae*-specific *apxIV* gene. The *apxIV*-positive sample was subjected to *de novo* assembly using Unicycler, and genome assembly quality was analysed using QUAST. Genome assemblies that passed the quality assessment were analysed using Kaptive against the KL database established in this study. The molecular type of the KL and details of the gene hit were reported by Kaptive and are shown in Tables S1 and S2.

(GenBank accession number MG780416.1) at the termini of the KL and the *cpxD* gene (GenBank accession number AIA09380) flanking the *cps* genes as previously described using Basic Local Alignment Search Tool (BLAST+) v2.7.1 [11, 21]. The full-length KL sequences and annotation information were obtained from the NCBI in GenBank file format (.gbk). By manually deleting the sequence of the mobile element IS*AplI* from the serovar 7 strain AP76 [22], a modified KL sequence with annotation was obtained, and this sequence was used as one of the KL reference sequences for serovar 7.

Due to the lack of a full-length serovar 6 KL sequence in the NCBI database, the serovar 6 strain Femo (ATCC 33590), a gift from the Animal Health Research Institute, Council of Agriculture, Executive Yuan, Taiwan, ROC, was subjected to WGS using both Illumina and Nanopore sequencing platforms. Briefly, for Illumina short-read sequencing, genomic DNA was prepared using the Gentra Puregene Yeast/Bact. kit (Qiagen), subjected to library preparation using the Nextera DNA Flex library prep kit (Illumina), and sequenced using an Illumina MiSeq sequencer. The short reads were trimmed using Trimmomatic v0.39 [23]. For Nanopore long-read sequencing, genomic DNA was prepared using the Quick-DNA HMW MagBead kit (Zymo Research), subjected to library preparation using a ligation sequencing kit (Oxford Nanopore Technologies; ONT), and sequenced using an ONT MinION sequencer. The long reads were trimmed by Guppy v4.0.14 (ONT) and filtered by Filtlong v0.2.0 [24]. The long-read data were subsampled and used to generate assemblies by genome assemblers, including Raven v1.1.10, Flye v2.8.1, Miniasm/Minipolish v0.1.3 and Redbean v2.5 [25–28]. The genome assemblies were used to generate a consensus genome using Trycycler v0.3.1 [29]. The consensus genome was polished with long reads by Medaka v1.0.3 (ONT; <https://github.com/nanoporetech/medaka>) and corrected with short reads by Pilon v1.23 [30]. The KL sequence in the genome of the *A. pleuropneumoniae* strain Femo was identified through analysis of the *modF*, *cpxD* and *ydeN* genes as mentioned above and annotated by Prokka v1.14.6 [31]. The KL sequence of the *A. pleuropneumoniae* serovar 6 strain Femo was deposited in the GenBank database with the accession number MZ450073.

A GenBank format file, including nucleotide sequences and annotations of coding sequences for reference KL types 1 to 19, was prepared with gene nomenclature consistent with that described by Bossé *et al.* [11]. Descriptions of KL types were added in the source feature of the GenBank file for compatibility with Kaptive (Supplementary data S2).

Table 1. Sequences used in the *A. pleuropneumoniae* KL database

KL type	GenBank accession no.	Serovar	Strain	No. of CDSs in the KL	Reference
1	CP029003.1	1	4074	14	[43]
2	ADXN01000030.1	2	4226	14	[44]
3	CP000687.1	3	JL03	14	[45]
4	LS483358.1	4	NCTC11384	11	[46]
5	CP000569.1	5	L20	13	[47]
6	MZ450073	6	ATCC 33590	16	This study
7-I*	CP001091.1	7	AP76	13	[48]
7†	CP001091.1†	7	AP76	12	This study
8	LN908249.1	8	MIDG2331	16	[49]
9	ADOI01000049.1	9	CVJ13261	14	[43]
10	ADOJ01000030.1	10	D13039	14	[43]
11	ADOK01000031.1	11	56153	13	[43]
12	ADOL01000042.1	12	1096	10	[43]
13	MG868947.1	13	N273	12	[11]
14	MG868948.1	14	3906	17	[11]
15	MG868949.1	15	HS143	11	[11]
16	MG868950.1	16	A-85/14	13	[11]
17	MG780416.1	17	16287-1	14	[11]
18	MG780423.1	18	7311555	11	[11]
19	MT468887.1	19	7213384-1	13	[4]

CDSs, coding sequences.

*Mobile element IS*Apl7* inserted in the KL.

†Sequence modified by deleting the mobile element in the KL reference sequence of KL07-I.

Collection and pre-processing of WGS short reads

Publicly available WGS reads of 189 *A. pleuropneumoniae* isolates and those of 66 samples from 14 non-*A. pleuropneumoniae* bacterial species derived from Illumina paired-end sequencing were obtained from the SRA database hosted at the NCBI. The SRA run accession numbers for WGS reads of the *A. pleuropneumoniae* and non-*A. pleuropneumoniae* bacterial species are listed in Tables 2 and S1, respectively. The SRA files were processed using SRA Toolkit v2.8.2-1 (NCBI) to obtain the raw WGS reads. For further analysis, the raw WGS reads were trimmed at a Phred quality score of Q30 and removed if they were shorter than 30 nucleotides using Trimmomatic v0.39 [23].

Screening of *apxIV* in short-read data

The presence of the *apxIV* gene in trimmed reads of *A. pleuropneumoniae* and non-*A. pleuropneumoniae* bacteria was analysed by ARIBA v2.14.6 [19], which mapped and assembled the reads with default settings against the *apxIV* dataset. Gene hits were reported when the alignment identity was $\geq 90\%$. The trimmed reads positive for *apxIV* were used for *de novo* genome assembly.

De novo genome assembly and quality assessment

The trimmed reads were *de novo* assembled using Unicycler v0.4.4 with an Illumina-only assembly pipeline [32]. Assembly statistics were analysed using QUAST v5.0.2 [33], and the quality of the genome assemblies was assessed. Genome assemblies with a minimum assembly length of 2 Mbp and a maximum contig number of 200 were subjected to manual inspection for the number and accumulative length of large contigs. Genome assemblies that passed the quality assessment were subjected to KL typing.

Table 2. WGS reads of non-*A. pleuropneumoniae* bacterial species used in this study

Species	No. of samples (n=66)	SRA run accession no.
Porcine bacterial pathogens	46	–
<i>Actinobacillus porcitosillarum</i> ^{*,*}	2	ERR200086, ERR200087
<i>Actinobacillus suis</i> *	4	SRR5184352, SRR5189134, SRR5189141, SRR5189321
<i>Bordetella bronchiseptica</i>	5	SRR931866, SRR942675, SRR9614213, SRR9614214, SRR9614215
<i>Escherichia coli</i> *	5	SRR10099931, SRR10099940, SRR10099946, SRR11647626, SRR9619978
<i>Glaesserella parasuis</i>	5	ERR175964, ERR176010, ERR176017, ERR225607, ERR270808
<i>Mycoplasma hyopneumoniae</i>	5	SRR7601664, SRR7601670, SRR7601671, SRR7601681, SRR7601682
<i>Pasteurella multocida</i>	5	SRR13148928, SRR13148936, SRR13148937, SRR13148946, SRR13148947
<i>Salmonella enterica</i> serovar Choleraesuis	5	ERR1777400, ERR1777414, ERR1777421, ERR1777449, ERR1777457
<i>Salmonella enterica</i> serovar Typhimurium	5	SRR2015698, SRR2015925, SRR2075991, SRR8291813, SRR9879548
<i>Streptococcus suis</i>	5	SRR4431635, SRR4431639, SRR4431646, SRR4431671, SRR5177695
Others	20	–
<i>Aggregatibacter actinomycetemcomitans</i> ^{*,†}	3	SRR12066794, SRR3170532, SRR3947678
<i>Bordetella pertussis</i> ^{*,†}	5	SRR11855995, SRR12105040, SRR5080696, SRR8689258, SRR9118293
<i>Mannheimia haemolytica</i> ^{*,‡}	5	SRR3749365, SRR3749458, SRR3750210, SRR3767545, SRR3775504
<i>Moraxella bovis</i> ^{*,‡}	2	SRR11012145, SRR7431214
<i>Proteus vulgaris</i> ^{*,†}	5	ERR4014498, ERR4014630, SRR10728100, SRR13191024, SRR13191127

*RTX toxin-producing bacteria.

†Human pathogens.

‡Bovine pathogens.

KL typing

The genome assemblies were analysed by Kaptive v0.7.3 against the *A. pleuropneumoniae* KL database with the default setting, and the best matching reference KL for each query genome assembly and a corresponding confidence level based on a BLASTN search and the number of genes were reported [15]. The match confidence was divided into six levels, namely, 'perfect', 'very high', 'high', 'good', 'low' and 'none', and the criteria for each level have been described elsewhere [14]. Except for the confidence level of none, all the other five levels were acceptable for serovar determination.

RESULTS

A. pleuropneumoniae KL database establishment

For KL typing of the genome assemblies, 20 KL sequences were used as reference sequences in the *A. pleuropneumoniae* KL database covering all known serovars 1 to 19 (Table 1). Each KL type was designated according to the serovar information of the reference sequences. To coordinate with Kaptive, each KL type was designated as KL and a number, e.g. KL01 to KL19, and a KL reference sequence was assigned to each KL type. The KL harbouring the insertion mobile element IS*AplI*, denoted 'KL07-I', was from *A. pleuropneumoniae* serovar 7 reference strain AP76 [22], while KL07 represented the modified reference sequence of KL07-I lacking IS*AplI*. The number of coding sequences in each KL ranged from 10 to 17 (Table 1). Based on a correspondence between KL type and serovar information of the reference sequence, the *A. pleuropneumoniae* serovar could be predicted accordingly, e.g. the query genome with a KL type of 'KLx' was predicted as 'serovar x'.

Screening of *apxIV* in WGS reads of *A. pleuropneumoniae* and non-*A. pleuropneumoniae* bacterial species

To assist with *A. pleuropneumoniae* typing and distinguish this species from other bacterial species, *apxIV* was used as an *A. pleuropneumoniae*-specific genetic marker [20]. The presence or absence of the *apxIV* gene in WGS reads of 189 *A. pleuropneumoniae* samples and those of 66 samples from 14 non-*A. pleuropneumoniae* bacterial species (Table 2) was screened using ARIBA against the *apxIV* dataset (Supplementary data S1). The presence of *apxIV* was detected in all the *A. pleuropneumoniae* samples tested (Fig. 2), indicating a sensitivity of 100% (189/189) using the WGS-based identification of *A. pleuropneumoniae*. Moreover,

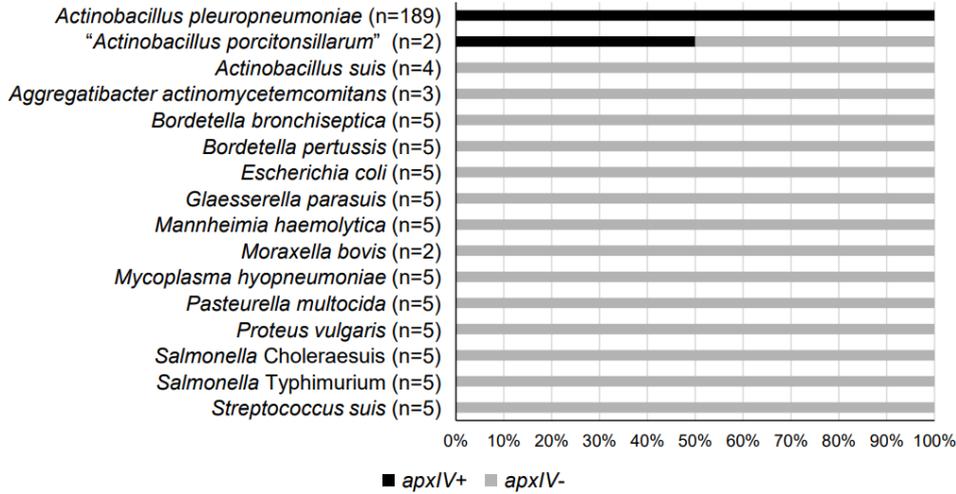


Fig. 2. Screening of *apxIV* using WGS short-read data. The trimmed WGS short-read data of 189 *A. pleuropneumoniae* and those from 14 non-*A. pleuropneumoniae* bacterial species were analysed by ARIBA against the *apxIV* dataset. The presence of *apxIV* was reported, and the percentages of *apxIV*-positive samples in each bacterial species were calculated.

the WGS reads from porcine, bovine and human bacterial pathogens, except for one “*Actinobacillus porcitonisillarum*” sample (SRA run accession number ERR200087), indicated that these organisms were *apxIV* negative (65/66). The *apxIV*-positive “*A. porcitonisillarum*” was subjected to further analysis. The results showed that the *apxIV* screening strategy distinguished the WGS reads of *A. pleuropneumoniae* from those of the non-*A. pleuropneumoniae* bacterial species.

De novo genome assembly of *A. pleuropneumoniae* WGS reads

After confirming the presence of *apxIV* in *A. pleuropneumoniae* short-read data, the trimmed reads were subjected to *de novo* genome assembly, and the assembly quality was assessed. A total of 105 of the 189 *A. pleuropneumoniae* assemblies passed quality assessment with a median contig number of 62 and a median assembly length of 2.26 Mbp. In this study, 32% (31/98) and 81% (74/91) of the genome assemblies derived from the paired-end 75 and 100 bp reads, respectively, passed quality assessment.

KL typing of the *A. pleuropneumoniae* genome assembly and interpretation

The 105 *A. pleuropneumoniae* genome assemblies were analysed by Kaptive analysis for KL type and serovar prediction. The results showed that the match confidence of 95% of the tested genomes (100/105) was good to perfect (Fig. 3a), indicating that a KL type was assigned to these genomes with high confidence. Thirteen KL types, including KL types 01, 02, 05–12, 15, 17 and 18,

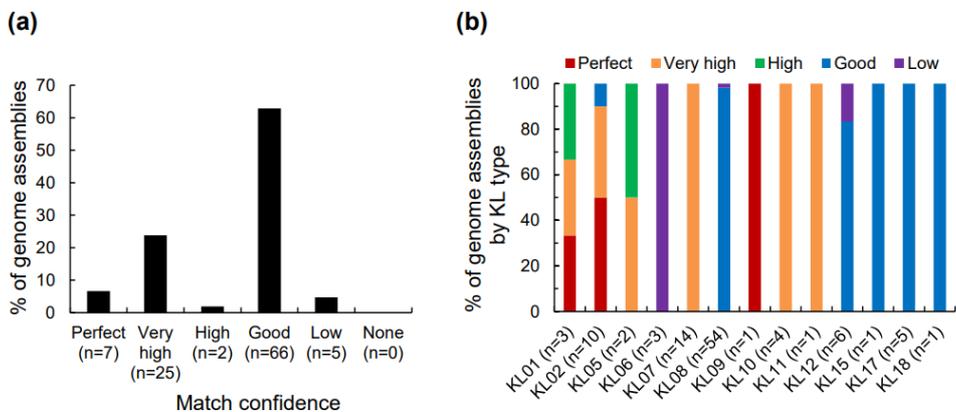


Fig. 3. *In silico* KL typing of *A. pleuropneumoniae* genome assemblies. The 105 *A. pleuropneumoniae* genome assemblies that passed the quality assessment were subjected to KL typing by Kaptive against the *A. pleuropneumoniae* KL database. The match confidence and KL types of the genome assemblies were reported by Kaptive. (a) Distribution of the match confidence of the assigned KL types of the 105 genomes. (b) Distribution of the match confidence by KL types.

Table 3. KL types and predicted serovars of the 105 *A. pleuropneumoniae* genome assemblies

KL type/predicted serovar	No. of test genomes	Serovar information from the literature (<i>n</i>) [reference]	Correspondence
1	3	1 (3) [50]	+
2	10	2 (9) [9, 50], K2:O7 (1) [50]	+
5	2	5 (2) [50]	+
6	3*	6 (2) [9, 50]	+
7	14	7 (14) [9, 50]	+
8	54*	8 (53) [9, 49, 50]	+
9	1	9 (1) [50]	+
10	4	10 (4) [50]	+
11	1	11 (1) [50]	+
12	6	12 (6) [50]	+
15	1	15 (1) [50]	+
17	5	ND (5) [50]	NA
18	1	ND (1) [50]	NA

+, A match between the serovar predicted based on KL type and that recorded in the literature.

*One genome lacking serovar information in the literature.

NA, not applicable; ND, not determined.

were assigned to the tested genomes (Fig. 3b). The most assigned KL types were KL08, with 53 good matches and a low match, and KL07, with 14 very high matches. The KL types of the 105 genome assemblies and the detailed results reported by Kaptive, including match confidence, coverage, identity, discrepancy in KL length and a full list of identified genes with identities, missing genes and extra genes, are shown in Tables 3 and S1.

We reviewed the Kaptive analysis results in detail and found that when a perfect or very high match was obtained, it represented an exact match or very close match to the KL reference sequence, e.g. the genomes of strains 405I and MIDG3457 (SRA run numbers ERR200067 and ERR200079) were typed as KL09 and KL11, respectively (Table S1).

Furthermore, we reviewed the gene hits in the Kaptive results (Table S1). The percentage of expected genes in the locus identified by Kaptive was at least 87%. Missing genes in the best-matched locus reported by Kaptive usually resulted from missing pieces of assembly or low gene identity, which was frequently accompanied by homologous genes with higher identity identified as other genes outside the best-matched locus. It was noted that there was a 256 bp deletion in the *cps2A* gene in the *A. pleuropneumoniae* strain MIDG3426 (SRA run accession number ERR200042), indicating the difference in KL sequences between the *A. pleuropneumoniae* strain MIDG3426 and KL02 reference sequence. As we observed, all of the five low matched samples, including three KL06 isolates, a KL08 isolate and a KL12 isolate, lacked *lysA* in the KL, which might be due to low-quality sequencing or sequence assembly of the KL, especially the region around the *lysA* gene.

Serovar prediction and confirmation

To examine the correspondence between KL type and serovar, we searched the serovar information of the WGS reads based on the SRA run accession numbers and strain names in the literature and compared them with the typing results from Kaptive. The predicted serovars of 97 tested genomes according to the assigned KL types, including KL01, KL02, KL05 to KL12, and KL15, were consistent with the serovar information documented in the literature (Table 3). The results showed an exact match (100%, 97/97) in the KL typing for *A. pleuropneumoniae* serovar prediction. Two genomes of strains MIDG2379 and Br384 (SRA run accession numbers ERR200009 and ERR200068) typed as KL06 and KL08, respectively, lacked serovar information in the literature. The genomes of *A. pleuropneumoniae* strains MIDG3419 to MIDG3422 and MIDG3440 (SRA run accession numbers ERR200035 to ERR200038 and ERR200050) and that of strain MIDG3435 (SRA run accession number ERR200049), of which serovars were previously undetermined, were typed as KL17 and KL18, respectively.

KL typing of the *apxIV*-positive “*A. porcitonisillarum*” strain MIDG3255

The WGS reads of the *apxIV*-positive “*A. porcitonisillarum*” strain MIDG3255 (SRA run accession number ERR200087) were *de novo* assembled and subjected to KL typing. The strain was classified as KL03 with a very high match. Moreover, all 14 genes in the KL were identified. *cps3F* shared 99.8% identity with the KL03 reference, and the remaining genes showed 100% identity with the reference. The length of the KL of the “*A. porcitonisillarum*” strain MIDG3255 was the same as that of the KL reference sequence of KL03. The detailed results reported by Kaptive are provided in Table S2. Furthermore, the G+C content of the genome assembly of “*A. porcitonisillarum*” strain MIDG3255 was 41.1%, which is more similar to that of *A. pleuropneumoniae* serovar 3 strain JL03 (41.2%; GenBank accession number CP000687.1) than that of “*A. porcitonisillarum*” strain 9953 L55 (39.7%; GenBank accession number CP029206.1). Collectively, the results show that the “*A. porcitonisillarum*” strain MIDG3255 is a misclassification of *A. pleuropneumoniae* serovar 3.

DISCUSSION

In the present study, we established an *in silico* analysis strategy to predict *A. pleuropneumoniae* serovars. This strategy combines *apxIV* screening of WGS short reads using ARIBA against the *apxIV* dataset to distinguish *A. pleuropneumoniae* and non-*A. pleuropneumoniae* bacterial species, and KL typing of genome assembly using Kaptive against the *A. pleuropneumoniae* KL database. This study is, to the best of our knowledge, the first to demonstrate that a genome-based approach is able to predict *A. pleuropneumoniae* serovars and to provide detailed information on the KL for further investigation.

The KL typing of *A. pleuropneumoniae* was highly accurate in this study, as we observed that Kaptive classified the 105 tested *A. pleuropneumoniae* genome assemblies into 13 KL types, including KL types 01, 02, 05–12, 15, 17 and 18. KL types 04, 13, 14, 16 and 19 were not identified in our study due to the unavailability of public WGS short-read data for serovar 4, 13, 14, 16 and 19 isolates. Nevertheless, the high discriminatory power of *in silico* KL typing was shown by the distinction between KL9 and KL11.

Among the 105 *A. pleuropneumoniae* genomes used for Kaptive analysis, the serovar information for 97 genomes can be found in the literature. Based on KL types, the predicted serovars of the 97 genomes were exactly the same as those recorded in the literature, indicating that *in silico* KL typing may serve as an alternative *A. pleuropneumoniae* typing method. Previously nontypable or potentially mistyped isolates may be classified through our *in silico* typing method, as exemplified that six nontypable *A. pleuropneumoniae* isolates were predicted as serovars 17 and 18. Furthermore, the *apxIV*-positive “*A. porcitonisillarum*” strain MIDG3255 was predicted to be *A. pleuropneumoniae* serovar 3. Analysis of the KL information was sufficient for serovar prediction without referring to gene loci associated with lipopolysaccharide biosynthesis. Nevertheless, several reports have indicated that some *A. pleuropneumoniae* isolates were designated as K:O serovars, including K1:O7, K2:O7, K19:O3 and K19:O4 [4, 34–36]. For further detection of K:O-serovar isolates of *A. pleuropneumoniae*, an expanding typing pipeline can be developed with a database of biosynthesis gene loci for the outer core of lipopolysaccharides. Collectively, *in silico* KL typing can be used in retrospective studies or routine molecular typing of *A. pleuropneumoniae* isolates. As reported in this and previous studies, Kaptive can be applied to genotyping or genomic characterization of *Acinetobacter baumannii*, *A. pleuropneumoniae*, *Klebsiella michiganensis*, *K. pneumoniae* and *K. variicola* [16–18, 37].

As Kaptive analysis provides detailed information on the KL, we recommend a manual reviewing of the result, including coverage, identity, discrepancy in KL length, and a full list of identified genes with identities and missing genes, to identify variants or potential new KLs and to understand the difference between the KL in the query genome and the KL reference sequence. In our study, lack of identification of *cps* genes was rarely accompanied by a length discrepancy exceeding 100 bp between the full-length KL extracted from the query genome and the reference sequence of the best-matched locus. If an exception occurred, it was worthy of further investigation, e.g. the KL in *A. pleuropneumoniae* strain MIDG3426 with a 256 bp deletion in the *cps2A* gene compared with the KL02 reference. This suggests that the KL in the strain MIDG3426 may be a potential new KL or a locus variant of serovar 2; however, more biological evidence is needed to confirm the variation in the antigenicity of CPS.

Although *in silico* typing has advantages, including high accuracy and potential for obtaining variant information in routine molecular typing and retrospective analysis of *A. pleuropneumoniae* isolates, it might have some limitations. One limitation is the dependence on the quality of WGS reads. WGS reads with appropriate read length, coverage and accuracy are the basis for *de novo* assembly [38]. Based on the SRA metadata of the *A. pleuropneumoniae* WGS reads analysed in this study, we suggested a minimum depth of 120× and 60× for paired-end 75 and 100 bp sequencing, respectively, and a genome size of 2.2 Mbp for *in silico* *A. pleuropneumoniae* typing. The contiguity and completeness of the genome assembly could also affect KL typing. A genome assembly with low contiguity might lead to discontinuous KL sequences in multiple contigs. Incompleteness of the genome might result from sequencing or assembly processes and lead to a poor match gene in Kaptive analysis. Low-quality genome assembly would lead to incorrect assignment of KL type; high-quality genome assembly would result in successful KL typing for serovar prediction.

During analysis, Kaptive evaluates any input genome, and the best-matched locus is given as the locus type for which the largest fraction of the locus has a BLAST hit against the assembly. In our unpublished data, by subjecting genome assemblies of

non-*A. pleuropneumoniae* bacterial species to Kaptive analysis, KL types were assigned even though the identity was extremely low, leading to incorrect interpretation of the KL type. Therefore, we suggest that identification of the *A. pleuropneumoniae*-specific *apxIV* gene is essential prior to *in silico* KL typing.

The species-specific *apxIV* gene is clinically important for *A. pleuropneumoniae* identification in molecular typing [20]. In the present genomic approach, *A. pleuropneumoniae* WGS reads were accurately distinguished from those of the other 14 bacterial species of porcine, bovine and human pathogens by screening for the presence of the *apxIV* gene in WGS reads by ARIBA. ARIBA was originally used to identify antimicrobial-resistance genes from short-read data in *Enterococcus faecium*, *Shigella sonnei* and *Neisseria gonorrhoeae* [19], and was applied to find the *A. pleuropneumoniae*-specific *apxIV* gene in our study. In a previous study, detection of *apxIV* with PCR could be used to distinguish *A. pleuropneumoniae* from other bacterial species, including *Actinobacillus equuli*, *Actinobacillus rossii*, *Actinobacillus suis*, *Actinobacillus minor*, *Actinobacillus porcinus*, *Haemophilus* spp. and *Mannheimia haemolytica* [20]. Furthermore, *Actinobacillus lignieresii* is closely affiliated with *A. pleuropneumoniae* [39], and whether *A. lignieresii* can be distinguished from *A. pleuropneumoniae* by our approach remains unknown. Because of the lack of *A. lignieresii* WGS data in short-read format in the NCBI SRA database, we were unable to examine *A. lignieresii* for the presence of *apxIV* by ARIBA, as ARIBA works on paired-end short reads. Alternatively, we analysed the genome assemblies of three *A. lignieresii* isolates (GenBank accession numbers GCA_900444945.1, GCA_900444935.1 and GCA_900635785.1) by BLAST. Compared to *apxIV* of *A. pleuropneumoniae*, the *A. lignieresii* genomes showed an identity of 70–72%, which was below the threshold (90%) required to identify the *apxIV* gene by the ARIBA tool. Therefore, we speculate that *A. lignieresii* would be distinguished as a non-*A. pleuropneumoniae* bacterium using our approach. Taken together, results showed that *A. pleuropneumoniae* could be distinguished from a total of 19 bacterial species by detection of *apxIV* in this and previous studies [20], demonstrating the consistent detection of *apxIV* as an *A. pleuropneumoniae*-specific marker. Nevertheless, identification of bacterial species could be achieved by analysis of the 16S ribosomal DNA (rDNA) gene [40, 41]. da Costa *et al.* reported that *A. pleuropneumoniae* was distinguished from *A. minor*, *A. porcinus* and *Pasteurella* spp. by 16S rDNA sequencing [42], indicating a potential use for 16S rDNA in the identification of *A. pleuropneumoniae* and phylogenetic analysis among closely related *Actinobacillus* species.

Collectively, the results of the present study show that *A. pleuropneumoniae* serovars can be predicted with a genomic approach through analyses of *apxIV* and KLs. *In silico* KL typing with Kaptive provides detailed genetic information and has advantages in the identification of potential variants or new KLs. As more KL sequences of new and regional prevalent serovars of *A. pleuropneumoniae* are obtained, the *A. pleuropneumoniae* KL database can be updated and expanded easily to ensure accurate assignment of KL types for *A. pleuropneumoniae* serovar prediction. This approach for *in silico* typing can be used for routine diagnosis, genomic surveillance and retrospective analysis of *A. pleuropneumoniae*.

Funding information

This study was supported by the Council of Agriculture, Executive Yuan, Taiwan, ROC, under grant numbers 109AS-8.8.1-BQ-B1 and 110AS-5.6.1-BQ-B1.

Acknowledgements

We thank the National Center for High-performance Computing (NCHC), Taiwan, ROC, for providing computational and storage resources.

Author contributions

Conceptualization: S.-C.L. and S.-L.H. Methodology: S.-C.L., J.-F.H., Z.-W.C. and S.-L.H. Formal analysis: S.-C.L. Investigation: S.-C.L. and J.-F.H. Validation: Y.-T.H. and H.-H.W. Resources: J.-P.W., J.-H.L., Z.-W.C. and S.-L.H. Data curation: S.-C.L., J.-F.H., Y.-T.H. and H.-H.W. Writing – original draft: S.-C.L. and S.-L.H. Writing – review and editing: S.-C.L., J.-F.H., Y.-T.H., H.-H.W., J.-P.W., J.-H.L., Z.-W.C. and S.-L.H. Visualization: S.-C.L. Supervision: S.-L.H. Project administration: Z.-W.C. and S.-L.H. Funding acquisition: J.-P.W., J.-H.L., Z.-W.C. and S.-L.H.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Shope RE. Porcine contagious pleuropneumonia. I. Experimental transmission, etiology, and pathology. *J Exp Med* 1964;119:357–368.
- Dubreuil JD, Jacques M, Mittal KR, Gottschalk M. *Actinobacillus pleuropneumoniae* surface polysaccharides: their role in diagnosis and immunogenicity. *Anim Health Res Rev* 2000;1:73–93.
- Perry MB, Altman E, Brisson J-R, Beynon LM, Richards JC. Structural characteristics of the antigenic capsular polysaccharides and lipopolysaccharides involved in the serological classification of *Actinobacillus (Haemophilus) pleuropneumoniae* strains. *Serodiagnosis Immunother Infect Dis* 1990;4:299–308.
- Stringer OW, Bossé JT, Lacouture S, Gottschalk M, Fodor L, *et al.* Proposal of *Actinobacillus pleuropneumoniae* serovar 19, and reformulation of previous multiplex PCRs for capsule-specific typing of all known serovars. *Vet Microbiol* 2021;255:109021.
- Ward CK, Inzana TJ. Identification and characterization of a DNA region involved in the export of capsular polysaccharide by *Actinobacillus pleuropneumoniae* serotype 5a. *Infect Immun* 1997;65:2491–2496.
- Ward CK, Lawrence ML, Veit HP, Inzana TJ. Cloning and mutagenesis of a serotype-specific DNA region involved in encapsulation and virulence of *Actinobacillus pleuropneumoniae* serotype 5a: concomitant expression of serotype 5a and 1 capsular polysaccharides in recombinant *A. pleuropneumoniae* serotype 1. *Infect Immun* 1998;66:3326–3336.
- Bossé JT, Li Y, Sárközi R, Gottschalk M, Angen Ø, *et al.* A unique capsule locus in the newly designated *Actinobacillus pleuropneumoniae* serovar 16 and development of a diagnostic PCR assay. *J Clin Microbiol* 2017;55:902–907.

8. Bossé JT, Li Y, Sárközi R, Fodor L, Lacouture S, et al. Proposal of serovars 17 and 18 of *Actinobacillus pleuropneumoniae* based on serological and genotypic analysis. *Vet Microbiol* 2018;217:1–6.
9. Bossé JT, Li Y, Rogers J, Fernandez Crespo R, Li Y, et al. Whole genome sequencing for surveillance of antimicrobial resistance in *Actinobacillus pleuropneumoniae* *Front Microbiol* 2017;8:311.
10. Birhanu BT, Lee SJ, Park NH, Song JB, Park SC. *In silico* analysis of putative drug and vaccine targets of the metabolic pathways of *Actinobacillus pleuropneumoniae* using a subtractive/comparative genomics approach. *J Vet Sci* 2018;19:188–199.
11. Bossé JT, Li Y, Fernandez Crespo R, Lacouture S, Gottschalk M, 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;220:83–89.
12. Turton JF, Perry C, Elgohari S, Hampton CV. PCR characterization and typing of *Klebsiella pneumoniae* using capsular type-specific, variable number tandem repeat and virulence gene targets. *J Med Microbiol* 2010;59:541–547.
13. Fang CT, Shih YJ, Cheong CM, Yi WC. Rapid and accurate determination of lipopolysaccharide O-antigen types in *Klebsiella pneumoniae* with a novel PCR-based O-genotyping method. *J Clin Microbiol* 2016;54:666–675.
14. Wick RR, Heinz E, Holt KE, Wyres KL. Kaptive Web: user-friendly capsule and lipopolysaccharide serotype prediction for *Klebsiella* genomes. *J Clin Microbiol* 2018;56:e00197–18.
15. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, et al. Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2016;2:e000102.
16. de Campos TA, de Almeida FM, de Almeida APC, Nakamura-Silva R, Oliveira-Silva M, et al. Multidrug-resistant (MDR) *Klebsiella variicola* strains isolated in a Brazilian hospital belong to new clones. *Front Microbiol* 2021;12:604031.
17. Klaper K, Hammerl JA, Rau J, Pfeifer Y, Werner G. Genome-based analysis of *Klebsiella* spp. isolates from animals and food products in Germany, 2013–2017. *Pathogens* 2021;10:573.
18. 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. *Microb Genom* 2020;6:e000339.
19. Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* 2017;3:e000131.
20. Schaller A, Kuhn R, Kuhnert P, Nicolet J, Anderson TJ, et al. Characterization of apxIVA, a new RTX determinant of *Actinobacillus pleuropneumoniae*. *Microbiology* 1999;145:2105–2116.
21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
22. Tegetmeyer HE, Jones SCP, Langford PR, Baltés N. ISAp1, a novel insertion element of *Actinobacillus pleuropneumoniae*, prevents ApxIV-based serological detection of serotype 7 strain AP76. *Vet Microbiol* 2008;128:342–353.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
24. Wick RR. rrwick/Filtlong: quality filtering tool for long reads; 2018. <https://github.com/rrwick/Filtlong>
25. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 2021;1:332–336.
26. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
27. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 2019;8:2138.
28. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 2020;17:155–158.
29. Wick RR. Rrwick/trycycler: Ttrycycler v0.3.1 (version v0.3.1). 2020.
30. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;9:e112963.
31. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
32. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
33. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
34. Gottschalk M, Lebrun A, Lacouture S, Harel J, Forget C, et al. Atypical *Actinobacillus pleuropneumoniae* isolates that share antigenic determinants with both serotypes 1 and 7. *J Vet Diagn Invest* 2000;12:444–449.
35. Nielsen R, Andresen LO, Plambeck T. Serological characterization of *Actinobacillus pleuropneumoniae* biotype 1 strains antigenically related to both serotypes 2 and 7. *Acta Vet Scand* 1996;37:327–336.
36. Kokotovic B, Angen Ø. Genetic diversity of *Actinobacillus pleuropneumoniae* assessed by amplified fragment length polymorphism analysis. *J Clin Microbiol* 2007;45:3921–3929.
37. Campos-Madueno EI, Sigrist T, Flückiger UM, Risch L, Bodmer T, et al. First report of a blaVIM-1 metallo-beta-lactamase-possessing *Klebsiella michiganensis*. *J Glob Antimicrob Resist* 2021;25:310–314.
38. Chaisson MJ, Brinza D, Pevzner PA. De novo fragment assembly with short mate-paired reads: does the read length matter. *Genome Res* 2009;19:336–346.
39. Dewhirst FE, Paster BJ, Olsen I, Fraser GJ. Phylogeny of 54 representative strains of species in the family Pasteurellaceae as determined by comparison of 16S rRNA sequences. *J Bacteriol* 1992;174:2002–2013.
40. Christensen JJ, Andresen K, Justesen T, Kemp M. Ribosomal DNA sequencing: experiences from use in the Danish National Reference Laboratory for Identification of Bacteria. *APMIS* 2005;113:621–628.
41. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 2019;10:5029.
42. da Costa MM, Klein CS, Balestrin R, Schrank A, Piffer IA, et al. Evaluation of PCR based on gene apxIVA associated with 16S rDNA sequencing for the identification of *Actinobacillus pleuropneumoniae* and related species. *Curr Microbiol* 2004;48:189–195.
43. Xu Z, Chen X, Li L, Li T, Wang S, et al. Comparative genomic characterization of *Actinobacillus pleuropneumoniae*. *J Bacteriol* 2010;192:5625–5636.
44. Zhan B, Angen Ø, Hedegaard J, Bendixen C, Panitz F. Draft genome sequences of *Actinobacillus pleuropneumoniae* serotypes 2 and 6. *J Bacteriol* 2010;192:5846–5847.
45. Xu Z, Zhou Y, Li L, Zhou R, Xiao S, et al. Genome biology of *Actinobacillus pleuropneumoniae* JL03, an isolate of serotype 3 prevalent in China. *PLoS One* 2008;3:e1450.
46. Doyle S. Chromosome 1 of *Actinobacillus pleuropneumoniae* strain NCTC11384 genome assembly; 2018. <https://www.ncbi.nlm.nih.gov/nucleotide/LS483358.1/>
47. Foote SJ, Bossé JT, Bouevitch AB, Langford PR, Young NM, et al. The complete genome sequence of *Actinobacillus pleuropneumoniae* L20 (serotype 5b). *J Bacteriol* 2008;190:1495–1496.
48. Linke B. Complete genome of *Actinobacillus pleuropneumoniae* serovar 7 strain AP76; 2008. <https://www.ncbi.nlm.nih.gov/nucleotide/CP001091.1>
49. Bossé JT, Chaudhuri RR, Li Y, Leanse LG, Fernandez Crespo R, et al. Complete genome sequence of MIDG2331, a genetically tractable serovar 8 clinical isolate of *Actinobacillus pleuropneumoniae*. *Genome Announc* 2016;4:e01667-15.
50. Bhuller RK. Comparative genome analyses to understand the population biology and virulence of the pig pathogen *Actinobacillus pleuropneumoniae*. PhD Thesis, Imperial College London, UK, 2017.