Molecular detection and phylogenetic analysis of pigeon circovirus in China during 2016-2019

Haoran Wang¹, Hui Gao¹, Zhiwen Jiang¹, Leibo Shi¹, Pengwei Zhao¹, Yanming Zhang¹, and Chengbao Wang¹

¹Northwest Agriculture and Forestry University

November 9, 2020

Abstract

The Pigeon circovirus (PiCV) which contains a circular single stranded DNA (approximately 2 kb) belongs to the genus Circovirus and the family Circoviridae. PiCV infections in pigeons (Columba livia) have been reported worldwide. Nowadays, pigeon racing is becoming increasingly popular and considered to be a national sport in China, and even, the greatest competitions of racing pigeons are stake place in China. However, there is no epidemiologic data on PiCV infections among racing pigeons in China. To trace the prevalence, genetic variation and evolution of PiCV in sick and healthy racing pigeons, 622 samples were collected from 11 provinces or municipalities of China from 2016 to 2019. Samples were tested by polymerase chain reaction. The results showed that the positive rate of PiCV was 19.3% (120/622) at the sample level; 59.0% (23/39) at the club level, suggesting that the virus was prevalent in Chinese racing pigeons. A sequence analysis revealed that the cap genes of the PiCV strains identified in our study display high genetic diversity and shared nucleotide homologies of 71.9%–100% and amino acid homologies of 71.7%–100%. 28 and 37 unique amino acid substitutions were observed among the cap proteins and rep proteins of our PiCV strains, respectively. Furthermore, two initiation codons (GTG and ATT) of cap gene were newly found. A cap-gene-based phylogenetic analysis showed that the strains in this study could be further divided into six groups (A, B, C, E, G, H and I) and some of our strains are closely related to worldwide strains from different types of pigeons. A large number of recombination events (31 events) were also detected in the PiCV genomes from Chinese racing pigeons. These findings suggest that PiCV strains circulating in China exhibits higher genetic diversity.
Summary

The *Pigeon circovirus* (PiCV) which contains a circular single stranded DNA (approximately 2 kb) belongs to the genus *Circovirus* and the family *Circoviridae*. PiCV infections in pigeons (*Columba livia*) have been reported worldwide. Nowadays, pigeon racing is becoming increasingly popular and considered to be a national sport in China, and even, the greatest competitions of racing pigeons are stake place in China. However, there is no epidemiologic data on PiCV infections among racing pigeons in China. To trace the prevalence, genetic variation and evolution of PiCV in sick and healthy racing pigeons, 622 samples were collected from 11 provinces or municipalities of China from 2016 to 2019. Samples were tested by polymerase chain reaction. The results showed that the positive rate of PiCV was 19.3% (120/622) at the sample level; 59.0% (23/39) at the club level, suggesting that the virus was prevalent in Chinese racing pigeons. A sequence analysis revealed that the *cap* genes of the PiCV strains identified in our study display high genetic diversity and shared nucleotide homologies of 71.9%–100% and amino acid homologies of 71.7%–100%. 28 and 37 unique amino acid substitutions were observed among the *cap* proteins and *rep* proteins of our PiCV strains, respectively. Furthermore, two initiation codons (GTG and ATT) of *cap* gene were newly found. A *cap*-gene-based phylogenetic analysis showed that the strains in this study could be further divided into six groups (A, B, C, E, G, H and I) and some of our strains are closely related to worldwide strains from different types of pigeons. A large number of recombination events (31 events) were also detected in the PiCV genomes from Chinese racing pigeons. These findings suggest that PiCV strains circulating in China exhibits higher genetic diversity.

KEYWORDS

genetic diversity, molecular epidemiology, phylogenetic analysis, *Pigeon circovirus*, sequence analysis

INTRODUCTION

Members of the *Circoviridae*, within the order *Circovirus*, contain a small, circular, non-enveloped, single stranded DNA genomes that range from 1.7 to 2.5 kb in size (Stenzel et al., 2017). The *Pigeon circovirus* (PiCV) or the columbid circovirus (CoCV), together with porcine circoviruses (PCV) types 1 and 2, psittacine beak and feather disease virus (BFDV), duck circovirus (DuCV), goose circovirus (GoCV), canary circovirus (CaCV), raven (RaCV), starling (StCV), swan (SwCV), finch and gull (FiCV and GuCV) circoviruses belongs to the genus *Circovirus* and the family *Circoviridae* (http://www.ictvonline.org/). PiCV is a small, circular, non-enveloped DNA virus that contains single-stranded DNA approximately 2.0 kb in size. The genome of PiCV forms two main open reading frames (ORFs). The ORF-V1, located on the virion sense strand, encodes a replication-associated protein (*rep*) and the ORF–C1, located on the complementary sense strand, encodes a capsid protein (*cap*) (Mankertz et al., 2000; Todd et al., 2001). In contrast to the gene forming ORF-V1, the gene forming ORF–C1 of PiCV had been demonstrated to be highly genetically diverse (Stenzel & Koncicki, 2017; Zhang et al., 2015). The other ORFs (ORF-C2, ORF-C3 and ORF-C4), located on the complementary sense strand, encode three PiCV proteins with unknown functions (Mankertz et al., 2000; Todd et al., 2008). The circovirus infection in pigeons was first diagnosed in 1993 in the USA (Woods et al., 1993) and had been considered to be strongly associated with young pigeon disease syndrome (YPDS) including weakened racing performance, weight loss, lethargy, anorexia, respiratory distress and diarrhea (Raue et al., 2005).

Various methods have been applied to detect PiCV infection in clinical specimens. The original diagnose methods including electron microscopy, histology, dot blot hybridization (DBH), and in situ hybridization (Yamamoto et al., 2015) are time consuming (Smyth et al., 2001; Soike et al., 2001; Todd et al., 2002; Tsai et al., 2014). The molecular biology techniques, such as the standard PCR, real-time PCR (Duchatel et al., 2009), NGS techniques (Wang et al., 2017) and the loop-mediated isothermal amplification method (Tsai et al., 2014), enabled the more rapid and accurate detection of PiCV infections. As a result, cases of PiCV
infections in pigeons were subsequently reported in various countries, including Northern Ireland (Todd et al., 2001), Germany (Mankertz et al., 2000; Raue et al., 2005), Italy (Franciosini et al., 2005), France (Abadie et al., 2001), Czech Republic (Taras et al., 2003), Belgium (Duchatel et al., 2005; Duchatel et al., 2006), Poland (Stenzel et al., 2012; Stenzel et al., 2014c), Slovenia (Krapez et al., 2012), Hungary (Cságora et al., 2012), United Arab Emirates (Ledwoń et al., 2011), Iran (Mahzounieh et al., 2014), China (Liao et al., 2015; Wang et al., 2017; Zhang et al., 2015), Japan (Yamamoto et al., 2015) and the USA (Roy et al., 2003). The previous studies also had demonstrated that PiCV has been detected in different types of pigeons including racing, fancy, feral, meat pigeons (Stenzel et al., 2014b; Stenzel et al., 2012; Wang et al., 2017). However, all of the mentioned above birds belong to *Columba livia* species.

In China, the first PiCV infection was detected from meat pigeons of Zhejiang province in 2009 and the full genome was sequenced (Yu et al., 2009). In recent years, several studies had proved that PiCV was prevalent among meat pigeons in eastern and southern of China (Wang et al., 2017; Zhang et al., 2015). The competitions of racing pigeons are becoming increasingly popular in China, and even, pigeon racing is considered to be a national sport. The Chinese Association of Racing Pigeon Breeders has over 5 million members. The Chinese breeders compete in numerous races at different distances in their sections through the racing season, and more than 25 million of racing pigeons were selected to attend the competitions in a total of approximately 750 racing clubs every year. However, there is no epidemiologic data on PiCV infections among racing pigeons. In order to investigate the prevalence, evolution and genome characterization of PiCV in racing pigeons in China, an extensive epidemiological investigation and bioinformatics analysis of PiCV from racing pigeons were undertaken. A total of 571 serum samples collected from healthy pigeons and 51 tissues samples collected from sick pigeons were obtained from 39 racing pigeon clubs in 11 provinces or municipalities of China between 2016 and 2019. The purpose of our study is to provide novel epidemiologic data and genome characterization about PiCV isolated from racing pigeons in China.

### 2 MATERIALS AND METHODS

#### 2.1 Ethics statement

This study was conducted according to the animal welfare guidelines of the World Organization for Animal Health, and animal sampling was carried out in strict accordance with guidelines established by the Ethics of Animal Experiments of Northwest A&F University, Yangling, China. All the protocols were approved by this committee (Permit Number: 2014BAD23B11). The field study did not involve endangered or protected species. No specific permissions were required for the collection of samples because the samples were collected from public areas or non-protected areas.

#### 2.2 Sample collection

In total, 571 serum samples were collected from apparently healthy pigeons from 39 racing pigeon clubs between November 2016 and August 2019 in 11 provinces or municipalities of China including Beijing, Hebei, Liaoning, Shanxi, Gansu, Qinghai, Shandong, Shaanxi, the Xinjiang Uygur Autonomous Region, Ningxia Hui Autonomous Region and Inner Mongolia Autonomous Region. Moreover, fifty-one fatally diseased racing pigeons from 7 racing pigeon club of the 5 provinces or municipalities were submitted to our laboratory in College of Veterinary Medicine, Northwest A&F University to determine the pathogen of the disease. The 0.2 g sections of internal organs (the heart, liver, spleen, lung, and intestine) were obtained during the postmortem examination of all mentioned above dead pigeons. All samples were stored at -80°C before DNA extraction.

#### 2.3 DNA extraction

The tissue samples were ground to powder with liquid nitrogen and diluted with three volumes of phosphate-buffered saline (PBS). The samples (1 ml) were centrifuged at 1,2000xg for 10 min at 4degC and the supernatants were transferred to a 1.5 ml tube. Nucleic acids from tissues and serum samples were extracted by using EasyPure Viral DNA/RNA Kit (TransGen Biotech, Beijing, China) according to the manufacturer’s instructions. The extracted genomic DNA was stored at -20degC before use.
2.4 PCR detection of PiCV

The PiCV was first detected using a PCR method targeting a 326-base fragment of *cap* gene described by Freick et al. (2008). The primer sequences were: PiCV-s, 5'-TTGAAAGGTTTTCAGCCTGGC-3' and PiCV-as, 5'-AGGAGACGAAGCACGCCTC-3' (Freick et al., 2008). The PiCV full genomes of all the positive samples were amplified by PCR using the primers: PiCV-1F, 5'-ACCCGCGACTTGGAGCCACGGAG-3' and PiCV-1R, 5'-TTGCCTCCGCATTGCGGTCGCT-3', PiCV-2F, 5'-GACACTAGTAAAGGGACCCAAGCCA-3' and PiCV-2R, 5'-AAGGCCTTGAGATGCGGGGT-3', respectively. PCR was performed by using Q5 Hot Start High-Fidelity 2X Master Mix (NEB, MA, USA) containing NEB Q5 Hot Start High-Fidelity DNA Polymerase (NEB, MA, USA). The contents of the three reactions mixture in a 50 μl reaction volume were as follows: 0.5 μM forward primer, 0.5 μM reverse primer, 1 μg of genomic DNA, 25 μl of Q5 High-Fidelity 2X Master Mix (NEB, MA, USA) and an appropriate volume of Nuclease-Free Water. The cycling parameters were 30 cycles of 98°C for 10 s, 55°C for 30 s and 72°C for 30 s, followed by a final extension at 72°C for 72 min using an automated BioRad T100 Thermal Cycler (Bio-Rad Laboratories, Inc., CA, USA). PCR products (5 μl) were resolved on 1% (w/v) agarose gels, stained with ethidium bromide, visualized with UV illumination inside a gel documentation apparatus (Bio-Rad Laboratories, Inc., CA, USA) and saved as digital photographs.

2.5 Genome characterization and sequence analysis of the PiCV genome and *cap* and *rep* genes of PiCV strains

The *cap* gene sequences (n=23), *rep* gene sequence (n=1) and the full genome sequences (n=67) of PiCV obtained in this study have been deposited in GenBank, respectively. The sequences obtained in the current study were compared with all available PiCV sequence data from different geographical locations within China (n=54) and the rest of the world (n=72) from the NCBI nucleotide database as reference sequences. Multiple sequence alignments were carried out using ClustalW within the MEGA 5.0 software (Tamura et al., 2011) and the homology among nucleotide and amino acid sequences was determined by using BioEdit v. 7.0.5 software (Hall, 1999). A divergence analysis of the *cap* protein (n=90) of PiCV was performed using WebLogo (http://weblogo.threeplusone.com/), an online software for sequence logo generator (Crooks et al., 2004). A comparison of entropy (Hx) in amino acid sequences of *cap* proteins (n=90) of PiCV was carried out by using GraphPad Prism 7.00 (GraphPad Software, Inc., USA).

2.6 Phylogenetic analysis

The entire *cap* gene sequences (n=90) of the PiCV obtained in this study were used for phylogenetic analysis. All available PiCV sequence data from different geographical locations within China (n=54) and the rest of the world (n=72) were retrieved from the NCBI nucleotide database as reference sequences. Maximum likelihood phylogenetic trees were constructed with the GTR+G+I model implemented in the MEGA 5.0 software, with partial deletion to handle alignment gaps, and 1000 bootstrap iterations. The phylogenetic tree was annotated with the Interactive Tree Of Life (iTOL) software (http://itol.embl.de/), an online tool for the display and annotation of phylogenetic trees (Letunic et al., 2006).

2.7 Recombination analysis

In order to analyze the potential recombinant strains, an integrated software package, the recombination detection program RDP4 (Martin et al., 2010), was used to detect potential recombinant strains, parental strains, and possible recombination breakpoints. Seven methods (RDP, GeneConv, BootScan, MaxChi, Chimaera, SiScan, and 3Seq) were implemented using the RDP4 program. Recombination events that were detected by at least three of the aforementioned methods with significant *p*-values (<0.05) were considered plausible recombinant events. Sequences in the analysed dataset that most closely resembled the parental sequences of recombinants were defined as either “minor parents” or “major parents” based on the size of the genome fragments that these sequences had contributed to the detected recombinants (with the major parent contributing the larger fragment and the minor parent the smaller).
3 RESULTS

3.1 PCR detection and prevalence analysis of PiCV infection in Chinese racing pigeons

In this study, PiCV was detected in samples from both the sicked and apparently healthy racing pigeons. Six hundred and twenty-two samples including serum samples (n=571) and tissue samples (n=51) obtained from 39 racing pigeon clubs in 11 provinces or municipalities of China between 2016 and 2019 were subjected to the detection the occurrence and prevalence of PiCV by using PCR technique. The results of PCR detection indicated that the positive rates of PiCV were 19.3% (120/622) at the sample level. Among the samples, 14 samples from the 51 sick pigeons (positive rate, 27.4%) and 106 samples from the 571 apparently healthy pigeons (positive rate, 18.6%) were PiCV positive, respectively. In our study, 23 out of the 39 racing pigeon clubs tested were positive for PiCV (positive rate, 59.0%) (Table 1). Detailed information on 120 PiCV positive samples from racing pigeons in China is shown in Table 1.

3.2 Amplification, sequencing and genome characterization of PiCV genomes

A total of Sixty-seven full genomes were successfully amplified, sequenced and assembled in this study were deposited in GenBank under Accession numbers: MW181925 to MW181991 (Table 2). The assembled circular whole genome sequences were detected as eleven multiple sizes and was ranging from 2030 to 2045 nt in sizes with their cap gene ranging from 813 to 828 nt in length and rep gene of 948 or 954 nt as shown in Table 2 and S1. Among these variable genome length, the most common genome sizes were encompassing 2037 (n=21), and 2042 (n=13) nucleotides, respectively (Table 2). The sequence comparison of the complete PiCV genomes revealed nucleotide homologies of 84.2%–100% among the 67 identified PiCV strains. The complete genome sequences of the 67 identified PiCV strains exhibited sequence similarities with Chinese PiCV reference strains (83.0%–97.8% nucleotide identity) and the PiCV reference strains from the other countries (82.6%–98.3% nucleotide identity) (Table 3). In particular, four unique nucleotide substitution (CG-GCGA) was found in the intergenic region between the rep and cap stop codons of two strains (SX1/SN/2017/MW181932 and YB4/SN/2018/MW181981) compared with the 67 PiCV strains obtained in this study and the other reference PiCV strains available from the NCBI GenBank, respectively.

3.3 Analysis of nucleotide sequence of cap gene and amino acid sequence of the cap protein

To explore the genetic relationship and evolution of PiCV from 2016 to 2019, the cap genes of 90 PiCV strains of the 120 PiCV positive samples were successfully sequenced. Detailed information on the 90 PiCV strains is shown in Table 2 and S1. The cap gene of 90 samples obtained in this study were used to compare with reference sequences from the Chinese and the other countries PiCV reference strains. The results showed that the cap gene of all 90 strains was ranging from 813 to 828 nt in length. Fourteen, thirteen, one, fifty-three, seven, and two of the 90 cap nucleotide sequences were 813 nt, 816 nt, 819 nt, 822 nt, 825 nt, and 828 nt in length, encoding a cap protein of 270, 271, 272, 273, 274, and 275 amino acid residues, respectively (Table 2 and S1). The sequence comparison of the cap genes revealed nucleotide homologies of 71.9%–100% and deduced amino acid homologies of 71.7%–100% among the 90 PiCV strains. The cap genes of the 90 identified PiCV strains exhibited low sequence similarities with Chinese PiCV reference strains (73.0%–99.6% nucleotide identity, 72.3%–100% amino acid identity) and the PiCV reference strains from the other countries (68.8%–98.4% nucleotide identity, 63.6%–100% amino acid identity) (Table 3).

Analysis of nucleotide sequences in the cap gene fragment indicated the appearance of many mutations. The strains identified in this study were formed into four groups (ATG (n=80), ATA (n=6), ATT (n=2), and GTG (n=2)) as the result of the mutation at the initiation codons (Table 2 and S1). To investigate variations in the deduced amino acid sequences of cap gene products, the amino acid sequences of 90 cap proteins and the reference strains were aligned. The results showed that there are nine major locations of deletions (compared to the consensus sequence) among the cap proteins. These include locations 7, 24, 29, 30, 35, 58, 130, 182, and 266 (Figure 1). A comparison of entropy (Hx) in amino acid sequences of cap proteins showed that the Hx of amino acid sequences at most positions from the 90 identified PiCV strains were higher than that of the PiCV reference strains (Figure 2). In addition, some unique amino acid substitutions at 28 different positions were observed among the cap proteins of 90 identified PiCV strains as
3.4 Analysis of nucleotide sequence in the rep and amino acid sequence of the rep protein

The rep genes of 68 PiCV strains of the 120 PiCV positive samples were successfully sequenced. Detailed information on the 68 PiCV strains is shown in Table 2 and S1. The rep gene of 68 samples obtained in this study were used to compare with PiCV reference sequences from China and other countries. The results showed that the rep gene of all 68 strains had been detected as two sizes: 948 nt and 954 nt. Forty of the 68 rep nucleotide sequences were 948 nt in length, encoding a rep protein of 315 amino acid residues. Twenty-eight of the 68 rep nucleotide sequences were 954 nt in size, encoding a rep protein of 317 amino acid residues (Table 2 and S1). The sequence comparison of the rep genes revealed nucleotide homologies of 90.3%–100% and deduced amino acid homologies of 92.7%–100% among the 68 PiCV strains. The rep genes of the 68 identified PiCV strains exhibited high sequence similarities with Chinese PiCV reference strains (89.0%–99.2% nucleotide identity, 89.2%–99.6% amino acid identity) and the PiCV reference strains from other countries (89.5%–98.3% nucleotide identity, 90.5%–99.3% amino acid identity) (Table 3). To investigate variations in the deduced amino acid sequences of rep gene products, the amino acid sequences of 68 PiCV strains including the reference strains were aligned. The results showed that some unique amino acid substitutions at 37 different positions were observed among the 68 identified PiCV strains as shown in Table 5.

3.5 Phylogenetic analysis of the cap gene

Phylogenetic analysis based on the cap gene were performed to investigate genetic relationship of the 90 PiCV strains obtained in our study with all the 126 PiCV strains reported in GenBank (Table 2, S1 and S2). The phylogenetic tree indicated that the total 216 analysed PiCV strains could be divided into 9 clades, which are defined by the contractual letters A–I (Figure 4). The 90 cap gene sequences reported in the present study were clustered into 7 clades: A (14.1%), B (60.9%), C (62.2%), E (25.0%), G (51.4%), H (28.6%) and I (40.0%). As shown in Figure 4, the PiCV strains from the different geographical locations clustered on the same groups, such as QYQX1/HE/2018/MW181909, TY1/SN/2016/MW181929, DS1/GS/2018/MW181970 and QD3/SN/2018/MW181965. On the other hand, the PiCV strains from the same geographical locations were divided into the different groups, such as LH1/HE/2018/MW181918 (clade B), LH2/HE/2019/MW181988 (clade A), and LH3/HE/2019/MW181989 (clade C). Furthermore, the PiCV strains collected in the different years were grouped closely in the same clade. Due to the fact that the racing pigeons were supplied by breeders from different provinces or municipalities of China, the 78 PiCV strains, isolated from the pigeons in the racing clubs in Shaanxi province, were belonged to six separate clades except for clade D, F and H.

3.6 Recombination analysis in PiCV genomes

For 67 identified in this study PiCV strains, 31 recombination events were detected using RDP4 (Table S3). These recombination events indicated that the strains originally from different types of pigeons or from different geographical locations may have some evidence of recombination. For example, a segment (1001bp) of three PiCV strains (LT3/SN/2018/MW181940, LT2/SN/2018/MW181939 and WQ6/SN/2018/MW181948) obtained in this study was likely descended from a large number of ancestral PiCV genomes that originated from Chinese meat pigeons, German racing pigeons, British racing pigeons, Polish fancy pigeons and French meat pigeons (event 2). The possible breakpoints for recombination were determined during the recombination analysis. The results also indicated that the recombination breakpoint hot plots were found to be located within both the intergenic region between the rep and cap stop codons, and near the virion strand origin of replication. Furthermore, the recombination breakpoint cold spots were detected within the central region of cap gene, which indicate little evidence for recombination within cap gene (Figure 5). Additionally, four PiCV genomes (TF1/SN/2016/MW181925, DA1/XJ/2018/MW181982, QD7/SN/2019/MW181990 and LH2/HE/2019/MW181988) had no evidence for recombination.

Discussion

shown in Table 4 and Figure 3.
In recent studies, RVA G18P[17] had been confirmed as a primary cause of YPDS-like diseases in domestic pigeons (Rubbenstroth et al., 2020). Furthermore, PiCV is also supposed to be an etiological agent of a multifactorial disease named young pigeon disease syndrome (YPDS) affecting mostly young pigeons worldwide (Rau et al., 2005). PiCV had been reported to be prevalent in at least fourteen countries. In China, PiCV has been first detected in meat pigeons since 2009 (Yu et al., 2009). In recent years, the epidemiology survey showed that the positive rates of PiCV infection in Chinese meat pigeons are 19.67% and 75.3% in the poultry farms of eastern China in 2009 and 2015 (Wang et al., 2017; Zhang et al., 2015), implying that PiCV is widely distributed among meat pigeon populations in eastern China. However, the epidemiology and distribution of PiCV infection in the racing pigeons is unknown. The main objective of this work was to evaluate the genetic diversity and epidemiology of PiCV strains circulating in the racing pigeons of China. Positive samples were detected from seven provinces and the prevalence rates among provinces are variant. The results showed that the positive rates of PiCV were 59.0% (23/39) at the club level and 19.3% (120/622) at the sample level. The data also indicated the positive rate in sick pigeons (positive rate, 27.4%) is higher than that of the apparently healthy pigeons (positive rate, 18.6%). Overall, our data for the first time implied that PiCV is also widely distributed in diseased racing pigeons and apparently healthy racing pigeons in China.

The previous studies had shown that the PiCV genome is 2031-2043 nucleotide in length and contains two major identified ORFs (ORF-V1 and ORF-C1) (Loiko et al., 2018). The most common genome sizes obtained in our study were encompassing 2037 (n=21), and 2042 (n=13) nucleotides, respectively. Furthermore, our findings first showed that the 2030 nt (n=1), 2044 nt (n=2), and 2045 nt (n=1) complete genome fragments were also existed in the PiCV positive samples. In addition, four unique nucleotide substitution (CG-GCGA) was first found in the intergenic region between the rep and cap stop codons of two strains (SX1/SN/2017/MW181932 and YB4/SN/2018/MW181981) compared with the 67 PiCV strains obtained in this study and the other reference PiCV strains. In this study, we also investigated whether PiCV strains isolated from a single club or from the same province were identical or similar to each other and evaluated the degree of their relatedness. Sequence analysis demonstrated that the sequence homologies between them were relatively low. The nucleotide sequence data revealed genetic diversity in the genome of the Chinese field PiCV strains, even though the viruses were obtained from the same place. Interestingly, two different isolates (QYQX1/HE/2018/MW181909, QYQX2/HE/2018/MW181910) with lower identity (75.5% nucleotide identity and 76.8% amino acid identity) in the cap gene were detected from the same pigeon. These results indicated that different strains were present in the same club. Though many genetic diversities were found in the genome of PiCV strains, no mutation was found in the conserved nonanucleotide motif (TAGTATTAC), located between the start codons of genes encoding rep and cap proteins. These data increase the understanding of the genetic diversities and the prevalence of PiCV in China. However, the pathogenicity and molecular characterization of the different PiCV still need to be validated in the future.

Previous studies have shown that there are many genetic diversities including insertions, deletions and substitutions in the genomes of PiCV, especially in cap gene (Cságló et al., 2012; Stenzel et al., 2014a; Wang et al., 2017). Our findings showed that the cap gene of the PiCV strains exhibits the highest diversity among themselves as well as in comparison with the reference strains. The cap gene of PiCV strains obtained in this study were also demonstrated to be highly genetically diverse with the length varied from 813 to 828 nt. In addition, the cap nucleotide sequences with 828 nt in length encoding for a novel cap protein of 275 amino acids was firstly been identified in two Chinese PiCV strains. Codons differing from ATG in a single position, collectively referred to as near-cognate initiation codons, are known to support translation initiation in eukaryotes (Wei, J. et al., 2013). As reported, two groups identified by ATG and ATA initiation codons of the cap gene had been found in the meat pigeons (Wang et al., 2017). In our study, four groups of initiation codons including ATG, ATA, ATT, and GTG were identified in the 90 PiCV strains among the racing pigeons in China. The previous results and our findings presented here showed that most of the PiCV strains were in the ATG group (Loiko et al., 2018; Todd et al., 2008; Wang et al., 2017). However, the ATT, and GTG group was firstly found in initiation codons of the cap gene in comparison with the reference strains in GenBank. Codons other than ATG had been identified to be less efficient initiation codons in
Neurospora (N.) crassa and human cells in vivo and in vitro (Wei, J. et al., 2013). Whether the different types of initiation codons of the cap gene have a hierarchy in initiation efficiencies need to be investigated in the future. Moreover, the cap gene of the current PiCV strains exhibits the highest diversity at the nucleotide and amino acid level and the rep gene is relatively conservative. Furthermore, 68 rep and 90 cap protein sequences were subjected to analysis of amino acid mutations, which show that there are numerous unique amino acid substitutions. These above-mentioned data combined with a comparison of entropy (Hx) in cap protein suggest that PiCV strains isolated from China had a higher level of diversity than the PiCV reference strains. These data also suggest that the Chinese PiCV strains may be undergoing more rapid mutational change in the racing pigeons. These data increase the understanding of the genetic diversity of PiCV. However, the pathogenicity and molecular characterization of the different PiCVs still need to be validated.

In the present study, a cap- gene-based phylogenetic analysis performed on the 90 cap genes obtained here and the other 126 cap genes available in GenBank revealed that the 90 PiCV strains were divided into seven clades (A, B, C, E, G, H, I). None of the 90 strains belonged to clades D and F, which contained the two Chinese meat pigeons and two Australia feral pigeons, respectively. These data suggested that the infection and evolution of PiCV in Chinese racing pigeons may have different evolutionary origins. This may result from the fact that the import of racing pigeons from all around the globe to China is very significant and the lack of oversight in the international pigeon trade (racing pigeons mainly) (Ashton, 1984; Stenzel et al., 2012). In addition, two novel PiCV strains (TY3/SN/2016/MW181931 and WL4/SN/2018/MW181959) with the ATT start codon and 828 nt in cap length formed a novel clade E with six Chinese PiCV strains from meat pigeons. The above indicates that some Chinese PiCV strains might evolved from the exclusively ancestors. On the other hand, the other two new identified in this study PiCV strains (TF5/SN/2016/MW181901 and WQ3/SN/2018/MW181945) with the GTG start codon form a small subclade with four European isolates in the clade A. Moreover, two different isolates (QYQX1/HE/2018/MW181909 and QYQX2/HE/2018/MW181910) obtained from one sick pigeon were clustered into different clades, which suggested horizontal transmission was occurred among the racing pigeons in the same racing clubs (Gerdes, 1993; Woods et al., 1993).

Viral recombination had been proved to play a significant role in the evolution of many ssDNA viruses (Lefeuvre et al., 2009). The extensive recombination had been reported in PiCV genomes and other circoviruses including PCV (Kleymann et al., 2020; Wei, C. et al., 2019) and BFDV (Julian et al., 2013; Varsani et al., 2011). Similar to the previously reports, a total of 31 recombination events were detected in the genomes of Chinese PiCV strains from the racing pigeons. Moreover, the recombination breakpoint hot plots were found to be located within both the intergenic region between the rep and cap stop codons, and near the virion strand origin of replication. The previous results and our findings presented here demonstrated that the recombination seems to be the key mechanism for the PiCV evolution (Csáhgola et al., 2012; Loiko et al., 2018; Sarker et al., 2019; Stenzel et al., 2014c).

In conclusion, our study demonstrated that PiCV infections in racing pigeons are widespread in the north of China and reveals characteristics of the PiCV genome. Furthermore, the identified PiCV strains displayed very high genetic diversity. This work has also demonstrated for the first time that PiCV from Chinese racing pigeons had extensive recombination. All the data increase our understanding of epidemiology and genetic variation of the PiCV circulating in China and evolutionary relationships amongst our strains with other worldwide strains. More molecular epidemiological information from other provinces in China and biological meaning of near-cognate codon initiation of cap gene is needed from future research.

ACKNOWLEDGMENTS

This study was funded by grants from Xi’an IRIS Livestock Technology Co. Ltd. (K4030218170), China Postdoctoral Science Foundation (2017M610659, 2018T111113), the State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences (SKLEV2016KFKT014), and the Fundamental Research Funds for the Central Universities (2452019053). The authors thank Ass. Prof. Tomasz Stenzel from Department of Bird Diseases, Faculty of Veterinary
Conflict of Interest Statement

The authors declare no conflict of interest.

References


**Figure legends:**

**Figure 1** Size and amino acid sequence differences of 90 cap proteins obtained in this study. Bars indicating the cap proteins are filled according to the phylogenetic group they occupy. The upper bar represents the consensus of the known PiCV cap proteins; Amino acids are numbered from 1 to 277. Gaps indicate the location of deleted (compared to the consensus sequence) amino acids. The number on the right is the length of cap proteins.

**Figure 2** Comparison of the amino acid sequence conservation at a particular position between the 90 identified PiCV strains and the PiCV reference strains available in GenBank. (a) The 90 identified PiCV strains from the racing pigeons in China. (b) The 54 Chinese PiCV reference strains available in GenBank. (c) The 72 PiCV reference strains available in GenBank from other countries. Note. The GenBank accession no. of the 90 identified PiCV strains, the 54 Chinese PiCV reference strains and 72 PiCV reference strains from other countries was shown in Table 2, S1 and S2.

**Figure 3** Divergence analysis of cap proteins of the 90 identified PiCV strains. (a) The 90 identified PiCV strains from the racing pigeons in China. (b) The 54 Chinese PiCV reference strains available in GenBank. (c) The 72 PiCV reference strains available in GenBank from other countries. The unique amino acid mutations found in this study are indicated in red. Note. All PiCV reference strains were the same to the Figure 2.

**Figure 4** Phylogenetic analysis of PiCV strains based on 90 cap gene sequences identified in this study and 126 cap gene sequences available in GenBank. A maximum likelihood phylogenetic tree was automatically constructed with 1000 bootstrap replicates and the remaining default parameters in the MEGA 5.0 software, and the Interactive Tree Of Life (iTOL) software was used for the display and annotation of the phylogenetic tree. Labels at branch tips refer to the strain name and GenBank accession number. Red taxa highlight the 90 PiCV genome sequences isolated from the racing pigeons geographically located in China.

**Figure 5** Recombination breakpoint distribution plots for 67 PiCV full genome sequences identified in this study and 113 PiCV full genome sequences available in GenBank. The red and blue areas of the plots indicate recombination breakpoint hot-spots and cold-spots, respectively. The dark and light grey areas represent the 95% and 99% confidence intervals of the expected degrees of breakpoint clustering under random recombination, respectively.

**Hosted file**


**Hosted file**
