Identification and Dynamics of DNA Viruses in Aves

by

Anthony Khalifeh

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2021 by the
Graduate Supervisory Committee:

Arvind Varsani, Chair
Simona Joop Kraberger
Greer Dolby

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

Viruses infect organisms in all domains of life and are abundant entities in ecosystems. In particular, single-stranded DNA viruses have been found in a wide variety of hosts and ecosystems. Using a metagenomic approach, novel circular viruses have been identified in multiple environmental samples. This thesis focuses on viruses and virus dynamics from avian sources. As part of this thesis, a novel phapecoctavirus was identified in a pigeon cloacal swab. The phapecoctavirus is most closely related to *Klebsiella* phage ZCKP1, identified from a freshwater sample. Beyond this, this thesis addresses circoviruses, which are of interest due to disease they cause to avian species. Evolution of circovirus recombination was studied in a closed system of uninfected and infected pigeons. 178 genomes of pigeon circovirus were sequenced, and patterns of recombination determined. Seven genotypes were present in the population and genotype 4 was shown to be present in a majority of samples after the experiment was finished. Circoviruses were also identified in waterfowl feces and the ten genomes recovered represent two new circovirus species. Overall, the research described in this thesis helped to gain a deeper understanding of the diversity and evolution of circular DNA viruses associated with avian species.

DEDICATION

Thanks to my family, my friends, and everyone who supported me, all of this work would not have been done if it were not for your constant support. For whether it be calling my family rambling about some scientific jargon or just telling them what I had for lunch that day, their support was much needed. To all my friends who would spend countless hours chatting whenever I had a problem or just playing games with me, thank you, I appreciate you for all the times you have spent hanging out with me.

ACKNOWLEDGMENTS

I would like to acknowledge the Varsani Lab team for their assistance and mentorship throughout my scientific journey. This thesis would not have been completed if it were not for the support that everyone gave to me. I would like to thank Dr. Arvind Varsani for being a great mentor and being always available to discuss my project. I would also like to acknowledge Dr. Simona Kraberger for always answering my questions, no matter how trivial they were as well as sharing every dessert with me. I would also like to acknowledge Rafaela Fontenele for giving me advice along the way and teaching me not to complain about doing anything that was not easy.

I would also like to acknowledge that multiple chapters of this work would not have happened if it were not for the blood and swab samples needed. I would like to thank Tomasz Stenzel and his lab for the collection of the samples and the experimental design.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

REVIEW

Viruses

Viruses infect organisms in all domains of life and are abundant entities in ecosystems. Viruses also cause diseases on certain organisms and a large body of work has focused on pathogenic viruses in the last five to eight decades. Developments in sequence-independent sequencing technologies have opened up the field of environmental virology and certainly have enabled the study of non-pathogen viruses in various animals. In general, the viruses are grouped based on their genome types in the Baltimore Classification system, i.e., double-stranded (ds) DNA, single-stranded (ss) DNA, reverse transcribing viruses with either DNA or RNA genomes, double-stranded (ds) RNA, negative sense single-stranded (ss) RNA, and positive sense ssRNA.

Classification of Circular DNA Viruses

A multitude of circular DNA (ss and ds) viruses have been identified in the past decade. This can be attributed to innovations in metagenomic sequencing and molecular biology. Circular DNA virus families that are associated with eukaryotes include *Anelloviridae, Ascoviridae, Bacilladnaviridae, Baculoviridae, Circoviridae, Geminiviridae, Genomoviridae, Hytrosaviridae, Marseilleviridae, Nanoviridae, Nimaviridae, Nudiviridae, Papillomaviridae, Polydnaviridae, Polyomaviridae, Redondoviridae, Redondoviridae, Smacoviridae* (Koonin et al., 2020). The advent of metagenomes has led to the discovery of many dsDNA viruses including bacteriophages (*Turriviridae, Ackermannviridae, Myoviridae, Siphoviridae, Podoviridae, Sphaerolipoviridae, Tectiviridae, Lipothrixviridae, Rudiviridae, Ampullaviridae,*

Globuloviridae, Tristromaviridae, Finnlakeviridae, Pleolipoviridae), focusing largely on those that infect *Salmonella* sp. and *Campylobacter* sp.

Classification of Circular Rep-encoding Single-stranded DNA Viruses

A multitude of ssDNA viruses have been identified in the past decade. Those with circular genomes which encode a replication-associated protein (Rep) are part of the phylum Cressdnaviricota (Krupovic et al., 2020). This group consists of 7 families of viruses including (*Bacilladnaviridae, Circoviridae, Geminiviridae, Genomoviridae, Nanoviridae, Redondoviridae, and Smacoviridae*) (Krupovic et al., 2020). Cressdnaviruses are associated with a variety of hosts and have been found in a wide range of environments (Table 1.1).

Family	Host	Virion ø, nm	Genome topology	Genome size, kb	ORFs	Replication protein[#]
Anelloviridae	animals	30–32	circular	2-4	2–4	Rep
Bidnaviridae	animals (insects)	20–24	linear	6–6.5 per segment (2 segments)	6	pPolB
Parvoviridae	animals	23–28	linear	4–6.3	3–4	HUH-Rep
Bacilladnaviridae	algae (diatoms)	33–38	circular	5.5–6	4	HUH-Rep
Circoviridae	animals	15–25	circular	1.7–2.1	2	HUH-Rep
Geminiviridae	plants	22x38	circular	2.5–3 per segment (1–2 segments)	4–8	HUH-Rep
Genomoviridae	fungi, insects	20–22	circular	2–2.4	2	HUH-Rep
Nanoviridae	plants	17–20	circular	0.98–1.1 per segment (6–9 segments)	5-8	HUH-Rep
Redondoviridae	?	?	circular	3.0–3.1	3	HUH-Rep
Smacoviridae	?	?	circular	2.3–2.9	2	HUH-Rep

Table 1.1: Families of ssDNA viruses associated with eukaryotic hosts. Families unified into the phylum Cressdnaviricota are in bold (Krupovic et al., 2020).

The *Circoviridae* Family

The *Circoviridae* family consists of two genera, *Circovirus* and *Cyclovirus.* *Cyclovirus* has been associated with a wide variety of hosts including dragonflies, human, bats, and rodents (Rosario et al., 2011; Smits et al., 2013). For the purpose of this study,

the genus *circovirus* was focused on as we investigate circovirus dynamics in avian species. Currently, as reported by ICTV (Walker et al., 2020), there are 43 species of circovirus, ten of which are associated to avian hosts, but this number is growing due to the 70 different species based upon GenBank uploads as of (2-12-21).

Genome Organization and Replication

Viruses in both genera have genomes that range from 1.7-2.1 kb. Members of the *Circovirus* genus encode a viral capsid protein (CP), replication-associated protein (Rep), and have an origin of replication (*ori*). The origin of replication is identifiable by a nanonucleotide ((T/n)A(G/t)TATTAC) sequence surrounded by an inverted repeat that forms a hairpin loop (Rosario et al., 2017). They replicate through the rolling-circle replication (RCR) mechanism and encode a Rep which encodes conserved motifs (Figure 1). These motifs are composed of two domains of amino acid sequences essential for replication, referred to as the rolling circle replication (RCR) motifs (Ilyina & Koonin, 1992). The HUH endonuclease domain consists of three amino acid motifs essential for DNA binding and cleavage (Chandler et al., 2013). The catalytic superfamily 3 helicase domain is composed of four motifs essential for DNA unwinding (Kazlauskas, Varsani, Koonin, & Krupovic, 2019). The genus *Circovirus* encodes the *ori* on the same strand as the *rep* ORF.

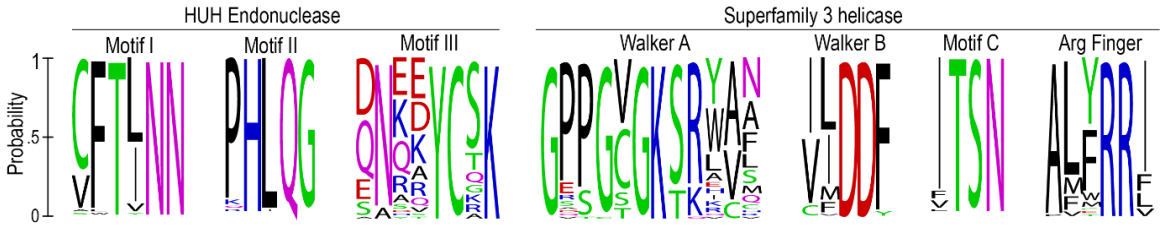


Figure 1.1: Frequency distribution of conserved sequence motifs of Rep protein. Residues are colored by chemical properties (polar, green; basic, blue; acidic, red; hydrophobic, black; neutral, purple).

Morphology

Using conventional electron microscopy (EM) along with advancements such as CryoEM, the structure of two circoviruses has been elucidated, porcine circovirus (PCV) and beak and feather disease virus (BFDV). Circoviruses form nonenveloped icosahedral virions that have T = 1 structure containing 60 identical protein subunits. These proteins consist of 12 flat pentameric morphological units (Crowther, Berriman, Curran, Allan, & Todd, 2003). Despite the similarity in protein number, BFDV is shown to have a more smooth, featureless virion compared to PCV virion structure. Virions of circovirus range from 12-32nm in diameter. BFDV has also been found to form 10nm decameric structures that are more prevalent when ssDNA was not present. This configuration also had the N-terminal ARM domain exposed, rather than interior in full virus like particles (Figure 2) (Nath et al., 2021).

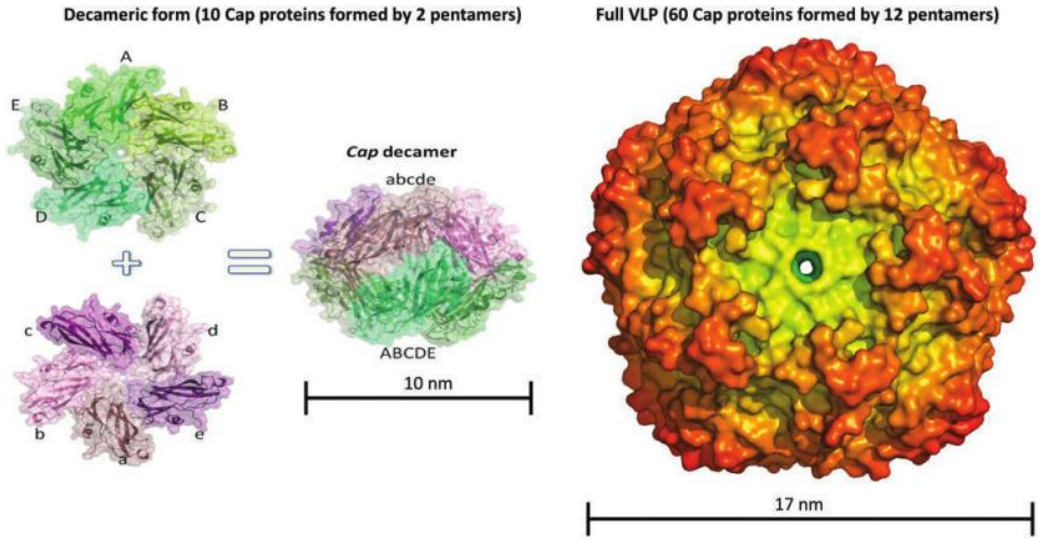


Figure 1.2: The beak and feather disease virus form at least two different macromolecular assemblies. A decameric species comprising two pentamers is ~10 nm in diameter. The full VLP comprises 60 capsid molecules (or 12 pentamers) and is 17 nm in diameter. (Nath et al., 2021)

Pathology

PCV and BFDV are the most extensively studied circoviruses as they both been shown to have a devastating impact on pig farming and parrot conservation, respectively. PCV2 has been shown to cause postweaning multisystemic wasting syndrome (PMWS), porcine dermatitis and nephropathy syndrome (PDNS), reproductive disorders and enteritis (Opriessnig, Meng, & Halbur, 2007; Segalés, 2012). BFDV is known to cause feather dystrophy, and causes beak and claw regrowth to have abnormalities (Pass & Perry, 1984; Schoemaker et al., 2000). Along with BFDV and PCV2, other circoviruses have negative health effects on their host including duck circovirus (DuCV), goose circovirus (GoCV), and pigeon circovirus (PiCV) (Daniel Todd, 2004).

Vaccination Efforts and Prevention

Due to many of the circoviruses economic, agricultural, and ecological impact (Bonne, Shearer, Sharp, Clark, & Raidal, 2009; Gillespie, Opriessnig, Meng, Pelzer, & Buechner-Maxwell, 2009; Y.-L. Huang et al., 2017; Zhaolong et al., 2020), there are multiple vaccines to lower and prevent infections of circovirus. PCV has multiple commercial vaccines that use different techniques including inactivated virus, capsid protein, and attenuated inactivated virus (Beach & Meng, 2012). BFDV has multiple vaccination efforts that are actively being researched and studied, including efforts using inactivated virus and recombinant capsid protein expression (Bonne et al., 2009; Raidal & Cross, 1994; Sarker, Ghorashi, et al., 2015). Duck circovirus vaccine efforts are in earlier

stages of development, with efforts focusing on inactivated virus and recombinant engineered capsid expression (Huang et al., 2018; Zhaolong et al., 2020). Pigeon Circovirus vaccine efforts have not been developed yet, but use of recombinant capsids to lower infection has been studied (T. Stenzel, Dziewulska, Tykałowski, et al., 2018).

Evolution

Circoviruses have varied evolutionary mechanisms that have been studied. Within the *Circovirus* family, there are significantly conserved sequences that are essential for replication and selection. Replication motifs are highly conserved throughout circoviruses as they are essential DNA binding, cleavage, and unwinding (Figure 1). Circoviruses appear to be highly host specific, with some studies showing signs of host-virus co-evolution (Figure 3) (Das et al., 2019; Johne, Fernandez-de-Luco, Höfle, & Müller, 2006). Circoviruses can evolve through recombination, a process that can occur through multiple pathways, including homologous recombination and genome rearrangement (Martin et al., 2011). Recombination has been shown to be an evolutionary force that effects multiple circoviruses including PCV and BFDV in a variety of hosts (Heath et al., 2004; Ma et al., 2007).

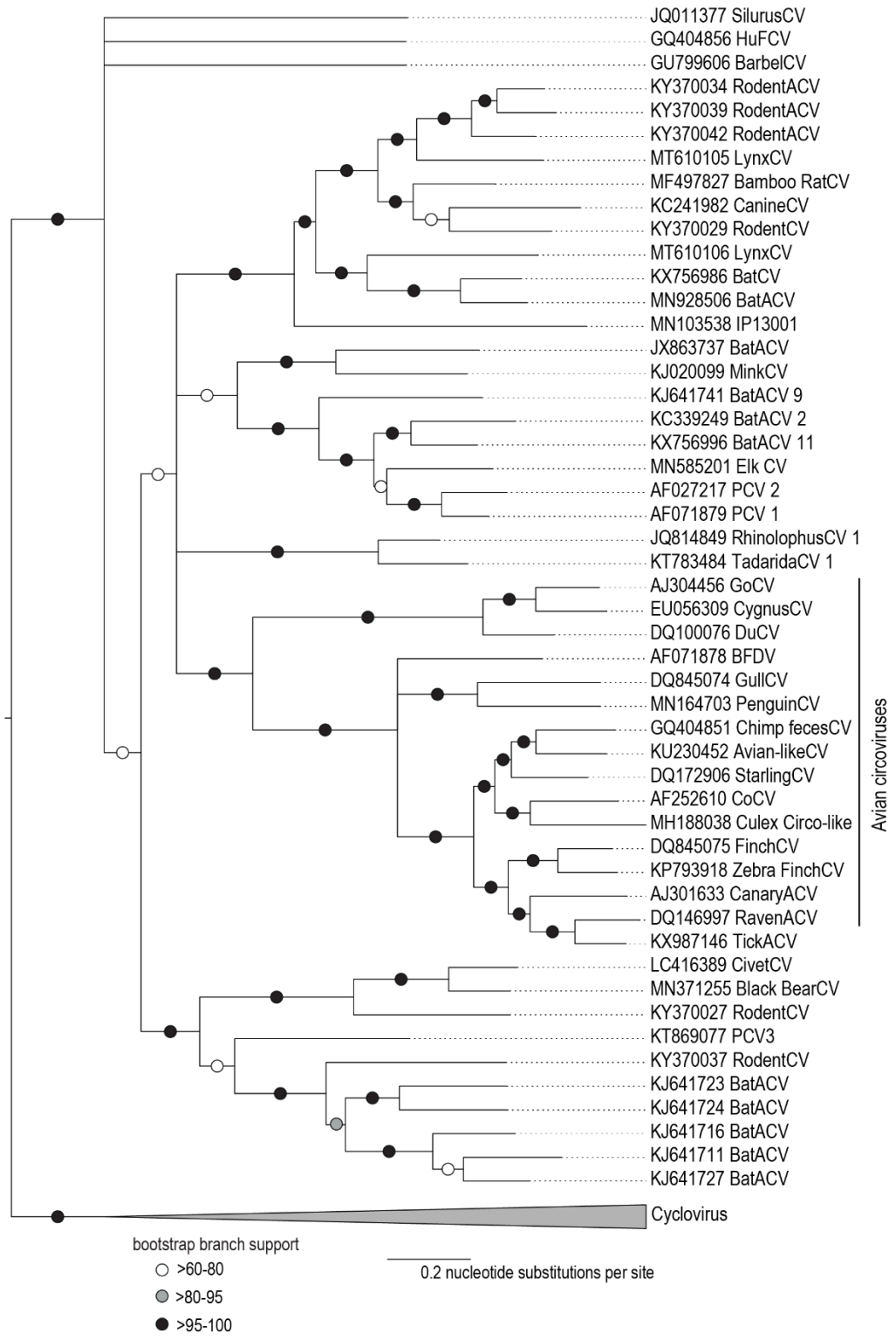


Figure 1.3: Maximum likelihood phylogenetic tree of aligned full genomes of representative circoviruses from each species grouping. Avian circoviruses are labelled.

Avian circoviruses

Pigeon circovirus

PiCV is found globally in both wild and domesticated pigeon populations (A. Stenzel, Pestka, Tykałowski, Śmiałek, & Koncicki, 2012; M. Coletti, M. Franciosini, G. Asdrubali, & F. Passamonti, 2000; Y.-L. Huang et al., 2017; Ledwoń, Bailey, O'DONOVAN, et al., 2011; M. Loiko et al., 2018; Leslie W Woods et al., 1994). Although PiCV has not been shown to directly cause any major disease in pigeons, a study has associated it with immunosuppression in younger pigeons (<4 months), which is a major contributing factor in young pigeon disease syndrome (YPDS) (Raue et al., 2005; Reitz, Kamphausen, Krautwald-Junghanns, Makert, & Schmidt, 2003). Studies show that pigeon circovirus can be transmitted both horizontally via fecal oral route and vertically (J.-P. Duchatel et al., 2005; J.-P. Duchatel, D. Todd, J. Smyth, J. Bustin, & H. Vindevogel, 2006; M. Franciosini et al., 2005). Phylogenetic analysis of PiCV does not show clear geographic structuring. Studies have shown pigeon circoviruses frequently recombine, specifically with a recombination hotspot within the intergenic region between *cp* and *rep* genes and near the origin of replication (Subir Sarker, Shubhagata Das, Seyed A Ghorashi, Jade K Forwood, & Shane R Raidal, 2019; T. Stenzel et al., 2014). Several approaches have been used to screen for PiCV including PCR, real time PCR, and LAMP assay (Duchatel, Todd, Willeman, & Losson, 2009; Tsai et al., 2014).

Beak and feather disease virus

BFDV has been found to infect wild and captive parrot species in Africa, Asia, Australasia, the Americas, and Europe. BFDV has been important due to disease outcome and ecological impact. BFDV causes Psittacine beak and feather disease (PBFD), which is

a potentially fatal disease. PBFD infects birds belonging to the Order Psittaciformes, of which 105 species are listed as endangered, vulnerable, or critically endangered (Raidal, Sarker, & Peters, 2015), but has shown evidence of spill-over into different orders of birds (Amery-Gale et al., 2017; Sarker, Lloyd, Forwood, & Raidal, 2016; Sarker, Moylan, et al., 2015).

Goose circovirus

GoCV was first isolated in Europe in 1999 from a farm-raised goose (Soike, Kohler, & Albrecht, 1999) and since has been detected in wild geese in Europe and farm raised geese in China and Taiwan (C-L Chen et al., 2003; T. Stenzel, Dziewulska, Muhire, et al., 2018; Yu, Zhu, Zheng, He, & Liu, 2007). GoCV displays clinical symptoms of immunosuppression, stunted growth, feather disorders, diarrhea (Guo et al., 2011; Soike et al., 1999). GoCV has been more studied in farmed goose and studies investigating its prevalence in wild geese are lacking.

Duck circovirus

DuCV was first identified in Germany in 2003 from a Mallard that belonged to a duck breeder (Hattermann, Schmitt, Soike, & Mankertz, 2003) and shortly after was identified in Taiwan (Chiou-Lin Chen et al., 2006). DuCV has been identified in Asia, Europe, and the Americas (Banda, Galloway-Haskins, Sandhu, & Schat, 2007). DuCV displays clinical symptoms of immunosuppression, feathering disorders, and low body weight. Although DuCV has been detected in wild migratory ducks in China (Niu, Liu, Han, Li, & Zeng, 2018), most studies revolve around farm-raised ducks.

Swan circovirus

Swan circovirus (SwCV) was first identified in naturally deceased mute swans (*Cygnus olor*) in Germany in 2006. Swan circovirus has only been sequenced in two swans. SwCV has not been shown to cause any symptoms that are associated with other avian circoviruses (Halami, Nieper, Müller, & Johne, 2008).

Other avian circoviruses

There are multiple other avian circoviruses, including canary circovirus, finch circovirus, gull circovirus, starling circovirus, and zebra finch circovirus (Johne et al., 2006; Phenix et al., 2001; Rinder, Schmitz, Peschel, & Korbel, 2015; Stewart, Perry, & Raidal, 2006; D. Todd, J. Weston, D. Soike, & J. Smyth, 2001). These circoviruses have been shown to infect native avian species but have not been linked to any disease. It is theorized that many of these circoviruses can cause similar symptoms as better studied circoviruses, e.g., pigeon circovirus and BFDV, but no conclusive studies have been done.

Approaches for the Discovery of Circular DNA Viruses

Rolling circular amplification

Rolling circular amplification (RCA) is a technique that utilizes the virus's natural replication cycle to produce copies of the genome. The RCA mechanism originates from the RCR mechanism, a process for ssDNA virus replication which has three stages: initiation, elongation, and termination (Gutierrez, 1999). RCA uses unique DNA polymerases that initiate in rolling circular replication (Mohsen & Kool, 2016), in particular bacteriophage phi29 DNA polymerase is used for efficient amplification of viral genomes. RCA allows for reliable replication of the viral genome without the need for expensive equipment (Haible, Kober, & Jeske, 2006).

PCR and qPCR-bases assays for detection of circoviruses

Polymerase chain reaction (PCR), is a technique designed to amplify DNA which uses multiple aspects of DNA replication to produce multiple copies of a specified region (Bartlett & Stirling, 2003). With specific primers, PCR can be utilized to isolate and scan for DNA of interest from a sample. This technology was expanded upon with the use of back-to-back primers. Back-to-back primers allow for amplification of the total genome of smaller size, therefore allowing downstream applications involving sequencing of the whole genome. Quantitative PCR, qPCR, is a technique that allows for real time quantification of the DNA quantity and the reaction using fluorescence (Wittwer, Herrmann, Moss, & Rasmussen, 1997)

High-throughput sequencing

High-throughput sequencing (HTS) is a sequencing technique is based on sequencing millions of short reads. HTS allows for detection of multiple viruses within a single DNA sample (Zhou et al., 2010). There are multiple platforms that utilize new techniques. HTS sequencing utilizes fragmenting the genome and then sequencing the fragments to allow for an abundance of small reads that are then assembled to form a more complete genome (Reuter, Spacek, & Snyder, 2015).

Specific Aims of This Study

The overarching theme of this thesis is to gain a better understanding of circular DNA viruses affecting avian species in both a closed system and environmentally.

CHAPTER 2

COMPLETE GENOME SEQUENCE OF A PHAPECOCTAVIRUS FROM A PIGEON CLOACAL SWAB SAMPLE

Abstract

The complete genome sequence of a bacteriophage in the genus *Phapecoctavirus* (family *Myoviridae*) from a domestic pigeon (*Columba livia f. domestica*) cloacal swab was identified using a high-throughput sequencing approach. The genome is 150,892 bp with a GC content of 39.1% containing 269 ORFs and 11 tRNA genes.

Introduction

Over the last decade, with the use of high-throughput sequencing, a plethora of known and novel viral sequences have been identified from a variety of sample types. For avian samples, cloacal swabs have been used for the identification of various pathogenic viruses such as circoviruses, flaviruses, gyroviruses, orthomyxoviruses, papillomaviruses, paramyxoviruses and polyomaviruses. Additionally, numerous bacteriophages have been identified including those classified in the *Myoviridae* family (Vibin et al., 2018). Myoviruses are dsDNA viruses that have contractile tails (Lavigne & Ceyssens, 2012). Identifying new viral genomes helps provide insights into viral diversity, evolution and putative hosts in the case of bacteriophages. Here we describe a new member of the *Phapecoctavirus* genus (family *Myoviridae*) identified in a cloacal swab collected from a 8-week old racing/ carrier domestic pigeon (*Columba livia f. domestica*) originated from amateur/ hobby pigeon facility located in Northern Poland (Samolubie, Bartoszyckie County, Warmińskpo-Mazurskie voivodeship, GPS: N54º11’02.1” E20º43’45.2”).

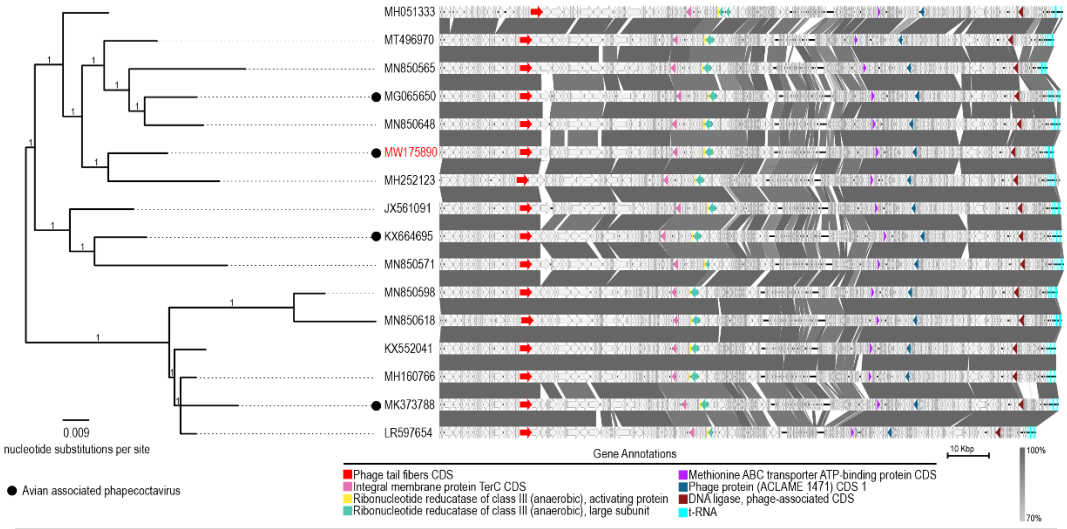
Materials and Methods

Viral DNA was extracted from pigeon cloacal swab using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA). Viral DNA was amplified using rolling circular amplification (RCA) with the TempliPhi 2000 kit (GE Healthcare, USA). The resulting RCA DNA was used to generate 2×150bp libraries at BGI (Hong Kong) using their DNBseq normal DNA library option and sequenced on their BGIseq sequencer. The resulting reads (11,610,789 read pairs) were quality trimmed using Trimmomatic v 0.39 (Bolger, Lohse, & Usadel, 2014) and the trimmed reads were *de novo* assembled using metaSPAdes v 3.12.0 (A. Bankevich et al., 2012). Contigs >1000nts were analyzed against a NCBI RefSeq (Brister, Ako-Adjei, Bao, & Blinkova, 2015) viral protein sequence database and bacteriophages were identified using Virsorter (Roux, Enault, Hurwitz, & Sullivan, 2015). All tools were run with default parameters unless otherwise specified. A circular contig (based on terminal redundancy) of 150,892nts (GC content, 39.1%) was identified that is most closely related to viruses in the genus *Phapecoctavirus* (family *Myoviridae*). 36072 reads mapped to this bacteriophage genome with a mean coverage of 16×. RASTtk (Brettin et al., 2015) was used to annotate this genome and predicted 269 open reading frames (varying in size of 96 to 3333 nts) and 11 tRNA genes (Arg, Asn, Gln, Gly, Ile, Met, Met, Pro, Ser, Thr, Tyr). We tentatively name this virus dompiswa virus (domestic pigeon swab associated virus). The genome sequence is deposited under the accession number MW175890 and mapped short reads are deposited in SRA SRR12914778.

Results and Discussion

All related genomes belonging to the genus *Phapecoctavirus* (n=15) were downloaded from GenBank (on the 12th of Nov 2020). The genomes were linearized at the

end of the cluster of tRNA genes, aligned with MAFFT (K. Katoh & D. M. Standley, 2013) and resulting alignment was used to infer a maximum likelihood phylogenetic tree with PHYML (Guindon, Delsuc, Dufayard, & Gascuel, 2009) with the WAG+G+I nucleotide substitution model (determined as best fit model using ModelTest (Posada & Crandall, 1998). Branches with <0.8 aLRT support were collapsed using TreeGraph2 (Stover & Muller, 2010) and midpoint rooted. Phylogenetically, dompiswa virus is most closely related to *Klebsiella* phage ZCKP1 (MH252123) isolated from freshwater (Taha, Connerton, Connerton, & El-Shibiny, 2018), sharing 92.3% inter-genomic distance determined using VIRIDIC (Moraru, Varsani, & Kropinski, 2020) and clusters with other unclassified phages (MH051333, MT496970, MN850565, MG065650, MN850648) (Figure 2.1).



Phapecoctavirus species	Name	Virus	Sequence length (nts)	Collection year	Country	Isolation source	Host /lab host
unclassified	MH051333	Escherichia phage vB_EcoM-Ro121c4YLVW	151671	2017	USA	liquid non-fecal compost	*Escherichia coli* STEC O121
unclassified	MT496970	Escherichia phage Mt1B1_P17	151202	-	France	sewage water	*Escherichia coli* Mt1B1
unclassified	MN850565	Escherichia phage ukendt	150947	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
● unclassified	MG065650	Campylobacter phage B15	150742	2016	USA	small farm chicken	*Campylobacter jejuni*
unclassified	MN850648	Escherichia phage anhysbys	149335	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
● unclassified	MW175890	Dompiswa virus	150952	2020	Poland	pigeon cloacal swab	-
Klebsiella virus ZCKP1	MH252123	Klebsiella phage ZCKP1	150925	-	Egypt	freshwater	*Klebsiella pneumoniae* KP/01
Escherichia virus phAPEC8	JX561091	Escherichia phage phAPEC8	147737	2010	Belgium	water sample in the vicinity of a poultry house	*Escherichia coli*
● *Escherichia virus ESCO5*	KX664695	Escherichia phage ESCO5	149312	2014	France	chicken fecal content	*Escherichia coli* BEN 5202
unclassified	MN850571	Escherichia phage nepoznato	151514	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
unclassified	MN850598	Escherichia phage nieznany	144998	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
unclassified	MN850618	Escherichia phage tuntematon	150473	2017	Denmark	wastewater	*Escherichia coli* K-12 MG1655
Escherichia phage ESCO13	KX552041	Escherichia phage ESCO13	149813	2014	France	water sample from a sewage treatment plant	*Escherichia coli* BEN 4311
unclassified	MH160766	Escherichia phage vB_EcoM-Ro121lw	149803	2017	USA	non-fecal compost	*Escherichia* sp. STEC O121
● *Escherichia virus Schickermooser*	MK373788	Escherichia phage vB_EcoM_Schickermooser	151194	2020	Germany	chicken feces	*Escherichia coli*
unclassified	LR597654	Escherichia phage phAPEC8_ev052	150152	2010	Denmark	human feces	-

Figure 2.1: Phylogenetic analysis of phapecoctavirus genomes (n=16). The accession number of dompiswa virus and associated information is highlighted in red. Genome comparison of phapecoctaviruses are shown to the right of the phylogeny. Gene annotations highlighted to show relative gene placement along genomes. Gray boxes between sequences indicate similarity of regions based upon BLASTn.

Conclusion

Five of the 16 phapecoctaviruses have been identified from avian fecal sources and using culture-based lab approaches, 14 have been shown to infect enterobacteria *Campylobacter jejuni*, *Escherichia coli* or *Klebsiella pneumoniae.* Thus, it is likely that dompiswa virus infects an enterobacteria but this needs to be confirmed.

CHAPTER 3

A PILOT STUDY INVESTIGATING THE DYNAMICS OF PIGEON CIRCOVIRUS RECOMBINATION IN DOMESTICATED PIGEONS HOUSED IN A SINGLE LOFT

Abstract

Pigeon circovirus (PiCV) is found in pigeon populations worldwide. It has been associated with immunosuppression in younger pigeons. Recombination is a mechanism of evolution that has previously been shown in various members of the *Circoviridae* family, including PiCV. In this study three groups of pigeons from separate lofts were screened for PiCV genotypes prior to being housed in a single one loft for a 3-weeks period. Post-grouping, blood and cloacal swab samples were taken for screening of PiCV genotypes with an aim to determine the recombination dynamics pre- and post-grouping the birds into a single loft. Genomes sequences of PiCV were determined from seven pigeons before they were housed together in a loft (n=58 sequences) and thereafter from all the ten pigeons either from blood or cloacal swabs (n=120). These 178 PiCV genome sequences group into 7 genotypes (98% genotype demarcation) and have an overall pairwise identity of 88-100%. Recombination analysis revealed 13 recombination events, and a recombination hot-spot spanning the 5' prime region the replication associated protein (*rep*) gene and the intergenic region. A cold-spot in the capsid protein coding region of the genome was also identified. The majority of the recombination region were identified in the *rep* coding region. This study provides insights into evolutionary dynamics regarding PiCV including the disease dynamics involving closed systems of pigeons rearing.

Introduction

Domestic pigeon (*Columba livia*), a species in the *Columbidae* family, are strongly associated with larger human populations (Haag-Wackernagel & Moch, 2004). Wild pigeons can flock in groups up to 400 individuals and have ranges that span several kilometers. Pigeons are also used extensively for racing and many studies have explored genetic, dietary, and infections on race performance (Freick, Muller, & Raue, 2008; Y. L. Huang et al., 2017; Omar, Hassan, & Shahin, 2017; Proskura, Kustosz, Dybus, & Lanckriet, 2015). During preparations for pigeon races and trainings, the practice of communal transporting, feeding, and watering birds from disparate lofts might create particularly ideal conditions for the transmission and long-distance dissemination of various pathogens. Multiple viruses have been associated and shown to infect pigeons, including those in the *Adenoviridae, Anelloviridae, Circoviridae, Herpesviridae, Paramyxoviruses*, *Parvoviridae*, *Picornaviridae and Reoviridae,* families (Agnihotri, Smith, Oakey, & Storie, 2021; Kaleta & Baldauf, 1988; McCowan et al., 2018; Phan et al., 2013; Teske et al., 2017; Wan et al., 2018; Zhang, Dai, & Dai, 2017).

Pigeon circovirus (PiCV, family *Circoviridae*), also referred to as columbid circovirus is a single-stranded circular DNA of ~2kb that encodes two bidirectionally transcribed genes, a replication associated protein (*rep*) and capsid protein (*cp*) genes and origin of replication (ori) (Mankertz, Hattermann, Ehlers, & Soike, 2000; D. Todd et al., 2008; D. Todd, J. H. Weston, D. Soike, & J. A. Smyth, 2001). PiCV is found globally in both wild and domesticated pigeon populations (M. Coletti, M. P. Franciosini, G. Asdrubali, & F. Passamonti, 2000; Y. L. Huang et al., 2017; Ledwoń, Bailey, O'Donovan, et al., 2011; M. R. Loiko et al., 2018; S. Sarker, S. Das, S. A. Ghorashi, J. K. Forwood, & S. R. Raidal, 2019; T. Stenzel & Pestka, 2014; T. A. Stenzel, Pestka, Tykalowski, Smialek,

& Koncicki, 2012; Wang, Zhuang, Qiu, Wang, & Chen, 2017; L. W. Woods et al., 1994). Although PiCV has not been shown to directly cause any major disease in pigeons, a study has associated it with immunosuppression in younger pigeons (<4 months) (Abadie et al., 2001; T. Stenzel, Dziewulska, Tykalowski, & Koncicki, 2020). Studies show that pigeon circovirus can be transmitted both horizontally via fecal oral route and vertically (J. P. Duchatel et al., 2005; J. P. Duchatel, D. Todd, J. A. Smyth, J. C. Bustin, & H. Vindevogel, 2006; M. P. Franciosini et al., 2005). Additionally, studies have shown pigeon circoviruses frequently recombine, specifically with a recombination hotspot within the intergenic region between *cp* and *rep* genes and near the origin of replication (S. Sarker et al., 2019; T. Stenzel et al., 2014). Positive selection appears to be stronger in the *cp* compared to the *rep* (S. Sarker et al., 2019; T. Stenzel et al., 2014). Phylogeny of PiCV sequences do not show any clear geographic structuring.

In recent years a popular pigeon sport called "One loft races" involving young pigeons has become popular. In this type of competition, all young birds supplied by various breeders are identically housed, fed, and trained before the race in one loft. However, such races violate all biosecurity principles and could lead to the transmission of contagious diseases. Bringing together divergent PiCV lineages and facilitating mixed infections of these, could also facilitate the new recombinant PiCV variants. This study shows the results of pilot investigation of PiCV recombination dynamics in a young, domesticated pigeons of different origin kept in the closed, common loft.

Materials and Methods

Ethical statement

The research protocol was approved by the Local Ethics Committee on Animal Experimentation of the University of Warmia and Mazury in Olsztyn (resolution No 41/2019, issued 28.05.2019, valid through: 01.10.2023). The researchers made every effort to minimize the suffering of birds.

Pigeon lofts and sample collection

8-week-old pigeons (n=10) were bought from 3 different pigeon lofts. Birds TSP1-4 originated from loft 1 (racing pigeons), TSP5-8 originated from the loft 2 (ornamental pigeons), TSP9-10 originated from the loft 3 (racing pigeons). Blood samples were collected on the 4th of May 2020 from the pigeons from the three lofts, subsequently, these pigeons were housed in a single loft. Four of the birds died during the experiment, so the second and final collection date was taken when they died, whereas the second date for the rest is the 26th of May 2020. Cloacal swabs were also collected from some of the birds randomly while in the common loft (Table 3.1, Figure 3.1).

Pigeon	Date	Blood	Cloacal swab	Accession #s
TSP1	04 May 2020	x		negative
	13 May 2020	x		MW656045, MW656046, MW656047, MW656048, MW656049
	06 May 2020		x	MW656103, MW656106, MW656107, MW656108, MW656125
TSP2	04 May 2020			negative
	26 May 2020	x		MW656050, MW656051, MW656052, MW656053, MW656054, MW656055, MW656056
	26 May 2020		x	MW656109, MW656110, MW656111, MW656104, MW656123
TSP3	04 May 2020	x		MW656022, MW656023, MW656024, MW656025, MW656026, MW656027, MW656028, MW656029, MW656030
	26 May 2020	x		MW656057, MW656058, MW656059, MW656060, MW656061, MW656062, MW656063, MW656064
	26 May 2020		x	MW656112, MW656113, MW656114, MW656122
TSP4	04 May 2020	x		negative
	07 May 2020	x		MW656065, MW656066, MW656067, MW656068, MW656069, MW656070, MW656071, MW656072, MW656073, MW656074
TSP5	04 May 2020	x		MW656128, MW656129, MW656130, MW656131, MW656132, MW656133, MW656134, MW656126, MW656135, MW656136
	06 May 2020		x	MW656105, MW656115, MW656116, MW656124, MW656127
	26 May 2020	x		MW655999, MW656079, MW656075, MW656076, MW656077, MW656078, MW656020, MW656144
	26 May 2020		x	MW656145, MW656149, MW656150, MW656151, MW656168
TSP6	04 May 2020	x		MW656153, MW656154, MW656159, MW656137, MW656155, MW656156, MW656157, MW656158
	06 May 2020		x	MW656162, MW656164, MW656165, MW656166
	26 May 2020	x		MW656080, MW656081, MW656082, MW656083, MW656084, MW656085, MW656086, MW656087, MW656088
	26 May 2020		x	MW656118, MW656119, MW656146, MW656169
TSP7	04 May 2020	x		MW656138, MW656143, MW656152, MW656139, MW656140, MW656031, MW656141, MW656120, MW656142
	26 May 2020	x		MW656089, MW656090, MW656091, MW656092, MW656093, MW656094
	26 May 2020		x	MW656147, MW656148, MW656117, MW656163, MW656167
TSP8	04 May 2020	x		MW655992, MW656032, MW656033, MW656034, MW656160, MW656161,
	15 May 2020	x		MW656000, MW656021, MW656095, MW656096, MW656097, MW656098, MW656099, MW656100, MW656101, MW656121
TSP9	04 May 2020	x		MW655993, MW655994, MW655995, MW655996, MW655997, MW656036, MW655998, MW656035
	06 May 2020		x	MW656015, MW656016, MW656017, MW656018, MW656019
	21 May 2020	x		MW656001, MW656002, MW656003, MW656004, MW656005, MW656006, MW656007
TSP10	04 May 2020	x		MW656037, MW656038, MW656039, MW656040, MW656041, MW656042, MW656043, MW656044
	26 May 2020	x		MW656008, MW656009, MW656010, MW656011, MW656102, MW656012, MW656013, MW656014

Table 3.1: Collection date and sampling data of PiCV isolates from study. Isolate accession numbers and number of isolates from samples listed.

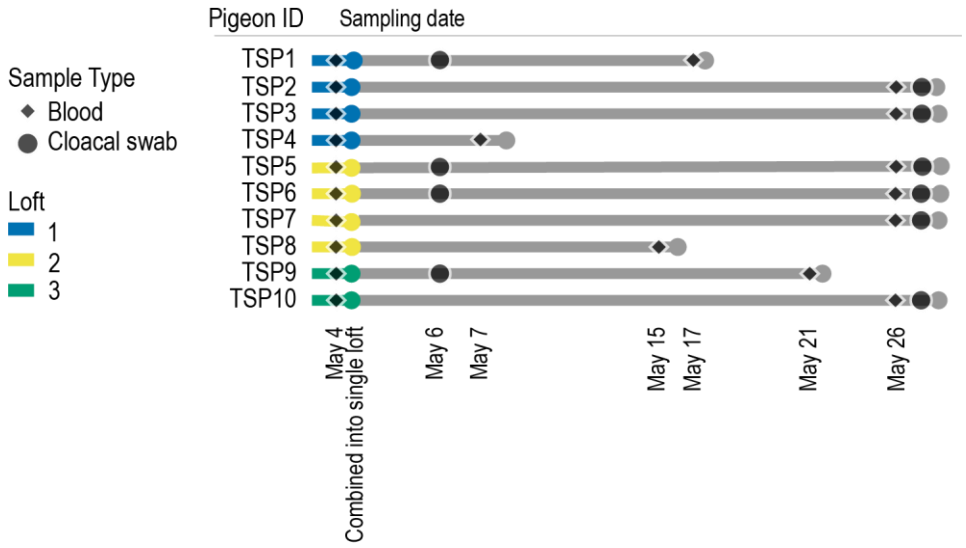


Figure 3.1: Summary of the timeline of blood and cloacal swab sampling dates of the pigeons for this study.

Viral DNA extraction and recovery of PiCV genomes

DNA was extracted from 10µl of blood samples using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) in accordance with the manufacturer's instructions. Viral DNA was isolated from 200 µl of UTM (Copan Diagnostics, Murrieta, CA) using the High Pure viral nucleic acid kit (Roche Diagnostics, USA) according to manufacturer's instructions. The DNA from the blood samples and cloacal swabs was amplified using rolling circle amplification (RCA) with TempliPhi 2000 kit (GE Healthcare, USA). Abutting primers (PICV_NF: 5'-VCG TGA CTT CAA AAC GGA AGT CAT C-3', PICV_NR: 5'-GGM TGC TGA CCA ATC AGC AGC TT-3') were designed in a common region in the genome based on an aligned PiCV dataset of all sequences available in GenBank. These were then used to amplify the PiCV sequences using polymerase chain reaction (PCR). The amplicons were resolved on a 0.7% agarose gel, ~2kb amplicons were excised from the gel and purified. The purified amplicons were ligated with the pJET 1.2 vector (Thermo Fisher Scientific, USA) and the recombinant plasmids were transformed into *Escherichia coli* XL

blue competent cells. Ten clones were selected from the blood and five from each swab positive sample for sequencing of the PiCV cloned genomes. The recombinant plasmids were purified using Fast DNA-spin Plasmid DNA Purification Kit (iNtRON Biotechnology, Korea) and Sanger sequenced at Macrogen Inc. (Korea) by primer walking. The reads were assembled and annotated using Geneious v11.0.3. 178 (Biomatters Ltd., New Zealand). After assembly, and removal of poorly sequenced clones, the number of sequences that were recovered from samples taken from each bird, pre and post movement of birds from three lofts into a single loft are summarized in Table 1.

Bioinformatic analyses of PiCV genome sequences

Full genomes from the PiCV genomes determined in this study together with those available in GenBank (downloaded on the 15th of Feb 2021) and beak and feather disease virus sequence as an outgroup were aligned using MUSCLE (Edgar, 2004). This alignment was used to infer a Neighbor-Joining tree using Mega X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018) with the Jukes-Cantor substitution model. Branch support with <60% bootstrap support were collapsed using TreeGraph2 (Stover & Muller, 2010). Genome-wide pairwise identities were determined using SDT 1.2 (Muhire, Varsani, & Martin, 2014).

Recombination analysis was performed using RDP5 (Martin et al., 2020) with default settings and only recombination events that were detected by more than three methods with a p-value <0.05 were accepted as credible.

The alignment of the sequences, with recombinant regions removed, was used to infer a Maximum Likelihood phylogenetic tree using MEGA (Kumar et al., 2018) using K2+G+I

nucleotide substitution model. Branches with <60% bootstrap support were collapsed using TreeGraph2 (Stover & Muller, 2010).

Results and Discussion

Characterization of PiCV genotypes

As part of a PiCV dynamics experiment, a total of 178 PiCV genomes were recovered from 10 birds from either their blood or cloacal swab samples (Figure 3.2). Ten birds were purchased from 3 different lofts and blood samples was collected from each of these animals. Thereafter they were all housed in a single loft for 22 days and blood and cloacal swab samples collect as summarized in Table 1 and Figure 2. A total of 58 genome sequences were recovered from 7 pigeons before they were housed together in a loft, and 120 thereafter (Table 3.1, Figure 3.2). From three of the pigeons (TSP1, TSP2 and TSP4) prior to grouping them in one loft, we did not find any PiCV in the birds (Table 3.1).

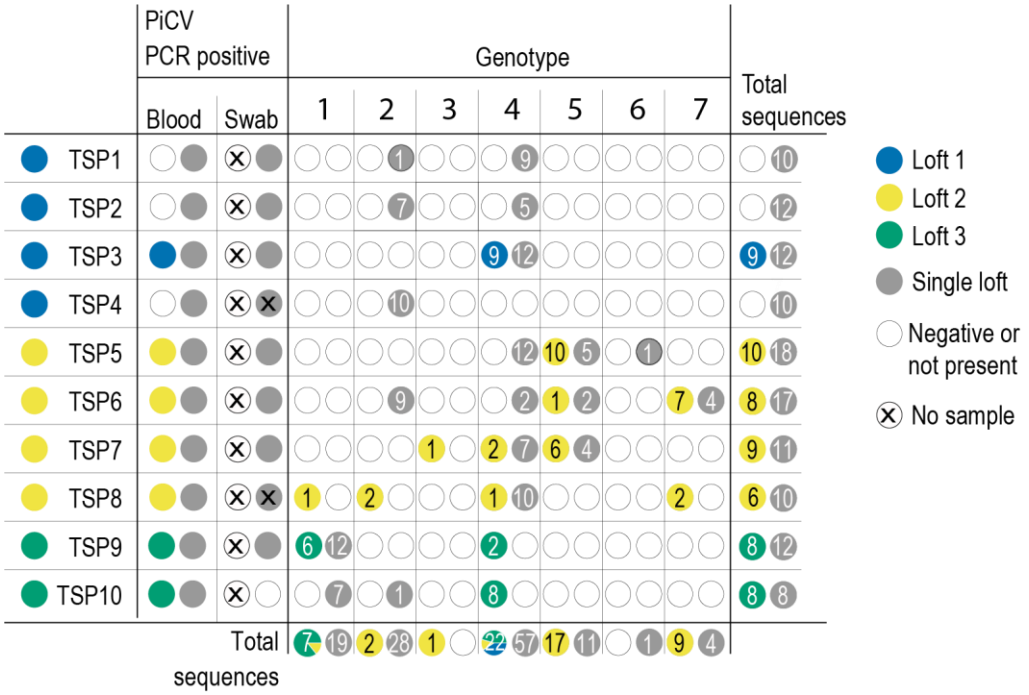


Figure 3.2: Summary of genotypes identified in the study in either the blood or cloacal swab samples. The pigeons were sourced from three separate lofts (loft 1-3) and then house in a single loft. Blood samples were taken when sourced and thereafter blood and cloacal swabs were taken at different time points as outlined in Figure 3.1. Genotypes identified in animals in each of the three lofts and also once they are housed in a single loft are color coded and the number in the circles indicates the number of genomes determined that are assigned to the genotype.

Based on the genome-wide pairwise identities, we were able to assign 7 genotypes to for all these PiCV genomes, with a genotype threshold of 98%. Genotypes 1, 2, 3, 4, 5, and 7 were present in the pigeons from the three lofts (Table 3.1, Figure 3.2). In one of the pigeons (TSP3, loft 1) had only genotype 4 it was identified. Once the pigeons from loft 1 (TSP1-4) were moved to a single loft, genotypes 2 and 4 were detected in TSP1 and -2, and genotype 2 in TSP4, suggesting that these animals became infected subsequent to being moved to a single loft. Alternatively, these genotypes were present in the animals but at a very basal level that they were not detected in our screening assays. In loft 2 animals, PiCV genotypes, 1, 2, 3, 4, 5 and 7 were detected at the initial stage of the experiment, whereas genotypes 2, 4, 5, 6 and 7 were present after housing the birds in the single loft, with TSP6 harboring four genotypes. PiCV genotype 3 was detected in TSP7 before all these animals were housed in a single loft, and this is the only animal to have this genotype. Moreover, PiCV genotype 6 was detected only in TSP5 pigeon once all of the birds were housed in a single loft. In the two animals in loft 3, PiCV genotypes 1 and 4 were detected, and once moved to a single loft, genotypes 1 and 2 were detected (Figure 3.2).

Genotype 4 was highly prevalent, making up 37.9% (22/58) sequences recovered from five pigeons in the three lofts. Once all the animals were housed in the same loft, this

genotype was found in seven pigeons account for 47.5% (57/120) of the recovered sequences (Figure 3.2).

These 178 PiCV sequences from this study were analyzed together with all PiCV full genomes available on GenBank (n=127). Pairwise identity analyses showed the PiCVs share ~83-100% similarity and based on a >98% nucleotide pairwise identity genotypes cut-off can be grouped into 102 genotypes. PiCV sequences from this study share >88% pairwise identity with each other representing 7 genotypes.

Phylogenetic analysis of PiCVs

PiCV genomes available in GenBank are from Australia (n=13), Belgium (n=6), Brazil (n=7), China (n=63), France (n=1), Japan (n=1), Nigeria (n=1), Poland (n=30), United Kingdom (n=2), United States of America (n=3). Global phylogenetic analysis of the PiCV recovered from 10 pigeons in this study from Poland (n=127) together with those PiCV genomes available in GenBank was undertaken (Figure 3.3). The phylogenetic analysis coupled with genotype assignment shows that PiCV sequences from this study are distributed within seven distinct clades which corroborate with the genotype designation based on >98% nucleotide pairwise identity threshold. Genotypes 3, 4, 5 and 6 form clades that sit most closely to PiCV sequences identified in a pigeon from Brazil and others from China and Poland. Genotypes 1 and 2 form clades with sequences that most closely related to PiCVs from birds sampled in Poland. Lastly, genotype 7 sequences cluster with two PiCV sequences from China. The PiCV genotypes identified in this study are distributed throughout the phylogenetic clearly highlighting the diverse PiCVs circulating in at least three of the lofts from which that the pigeons were acquired. Furthermore, we detected multiple genotypes within a single animal.

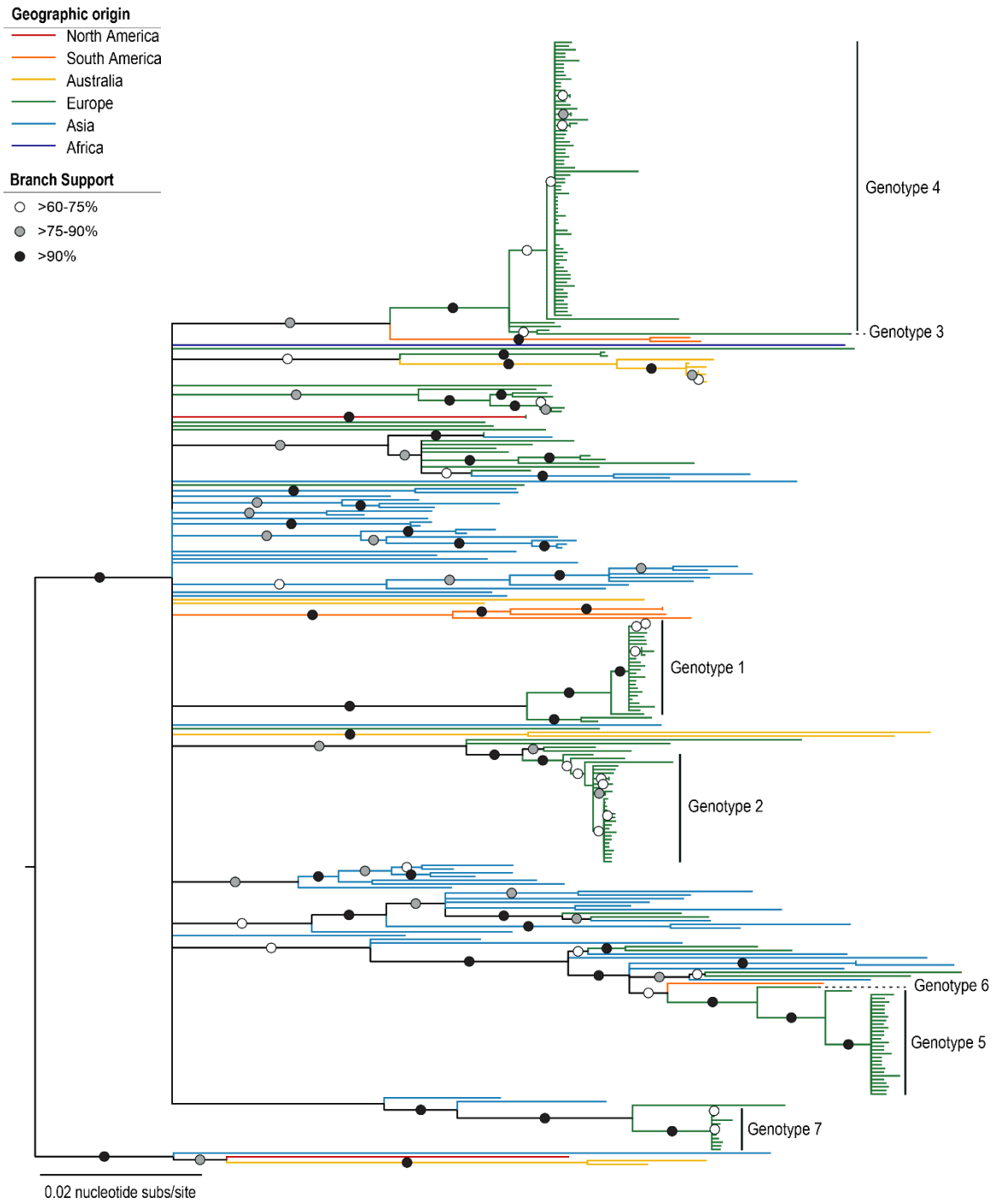


Figure 3.3: Neighbor-Joining phylogenetic tree of the PiCV sequences from this study together with those from GenBank. Branches are color coded based on geographical sampling of the PiCV.

Recombination analysis

To investigate PiCV recombination dynamics, a simulation of "housing together" naturally infected and non-infected animals originated from different sources was performed. The recombination analysis showed that there are 13 detectable recombination events. The recombination events detected in a large portion of the sequences was event 2, occurring in 109 genomes (Figure 3.4) that spans a ~100 nt region of the 5' end of *rep*. 12 of the recombination events are within *rep* gene (Figure 3.4). We identified a recombination hotspot spanning the 3' region of *rep* and the intergenic region whereas a cold spot was identified near the 5' region of the *cp* (Figure 3.5). These results differ to previous hot-spot detected in PiCV (T. Stenzel et al., 2014), which shows hot spots towards the 5' end of the *rep* and *cp*, as well as the intergenic region. The differences in recombination patterns could be due the variation in samples i.e., global datasets with "circulating" recombinants as opposed to the ones in this study that addresses recombination in a short time frame within a small set of animals in a loft. The cold spot located in the *cp* along with low amounts of recombination events show a higher region of conservation and importance of the *cp* likely for transmission and infectivity. Phylogenetic analysis of the genomes without recombination shows five clades present in the population (Figure 3.4) in contrast to the 7 genotypes present in the Neighbor-Joining phylogenetic analysis (Figure 3.3).

There is no evidence of recombination in sequences present in Clade II and IV (Figure 3.4). Clade IV (Figure 3.4) is mainly comprised of sequences determined from animals only prior to being in a single loft and those after being in a loft for a week. PiCV sequence MW656152 from animal TSP7 had 3 recombinant regions accounting for 42%

of the genome but this was from a sample taken from the animal while in maternal loft 2 and this represents PiCV recombination within the loft it was housed in when purchased (Figure 3.4; Table 3.1; Table 3.2). PiCV sequence MW656144 from TSP5 has a recombinant region, accounting for 34% of the genome. Both PiCV sequences MW656152 and MW656144 are singletons forming their own genotypes 3 and 6, respectively.

Only 3 events were identified in sequences derived from the animals while in the three separate lofts or origin whereas 10 recombination events were detected in sequences derived from animals once they were housed in the same loft.

Event #r	Begin	End	Recombinant sequence(s)	Minor parental sequence(s)	Major parental sequence(s)	Method	p-value
1	983	1512	Genotype 3	Genotype 6 (all) Genotype 5 (MW656131, MW65613, MW656150, MW656169)	Genotype 4 (all)	GBMCS**T**	8.88E-47
2	38	140	Genotype 4 (all) Genotype 2 (all) Genotype 3	Genotype 5 (all)	Genotype 1 (all)	**G**BMCS	6.04E-20
3	142	418	Genotype 2 (MW656088)	Genotype 7 (all) Genotype 5 (MW656159)	Genotype 2 (expect MW656088)	**G**MCST	2.77E-17
4	275	983	Genotype 6	Genotype 4 (MW656020)	Genotype 5 (all)	**G**MCST	4.69E-16
5	577*	816	Genotype 4 (MW656076) Genotype 3 (MW656152)[P]	Genotype 5 (all)	Genotype 4 (MW656020, MW656093, MW656106, MW656109, MW656120)	**G**MCST	1.52E-13
6	2018	158	Genotype 4 (MW656020)	Genotype 1 (all)	Genotype 2 (all)	**G**MCST	2.46E-13
6	2018	158	Genotype 4 (MW656020)	Genotype 1 (all)	Genotype 2 (all)	**G**MCST	2.46E-13
7	484	904	Genotype 2 (MW656047)	Genotype 4 (MW656020)	Genotype 2 (MW656085)	G**B**MCST	2.59E-13
8	644	856	Genotype 2 (MW656085)	Genotype 7 (all)	Genotype 2 (except MW656085)	**G**BMCST	4.99E-12
9	577*	948	Genotype 4 (MW656078)	Genotype 1 (all)	Genotype 4 (except MW656078)	**G**BMCST	5.69E-11
10	142*	508	Genotype 4 (MW656109)	Genotype 5 (all)	Genotype 4 (MW656093, MW656119, MW656020) Genotype 3	GBMCS**T**	8.88E-22
11	277	676	Genotype 5 (MW656169)	Genotype 4 (all)	Genotype 5 (except MW656169)	G**B**MCST	3.20E-17
12	624	752	Genotype 4 (MW656119)	Genotype 5 (all)	Genotype 4 (except MW656119) Genotype 2 (all)	G**B**T	2.32E-09
13	621	813	Genotype 1 (MW656017)	Genotype 5 (all)	Genotype 1 (except MW656017)	G**B**T	1.29E-08

Table 3.2: Summary of the recombination events detected in the genomes of PiCV from this study. The methods used to detect recombination are RDP (R), GENCONV (G), BOOTSCAN (B), MAXCHI (M), CHIMERA (C), SISCAN (S) and 3SEQ (T). For each recombination event, the method with the highest p-value is in bold font and underlined. [P] denotes partial evidence of recombination.

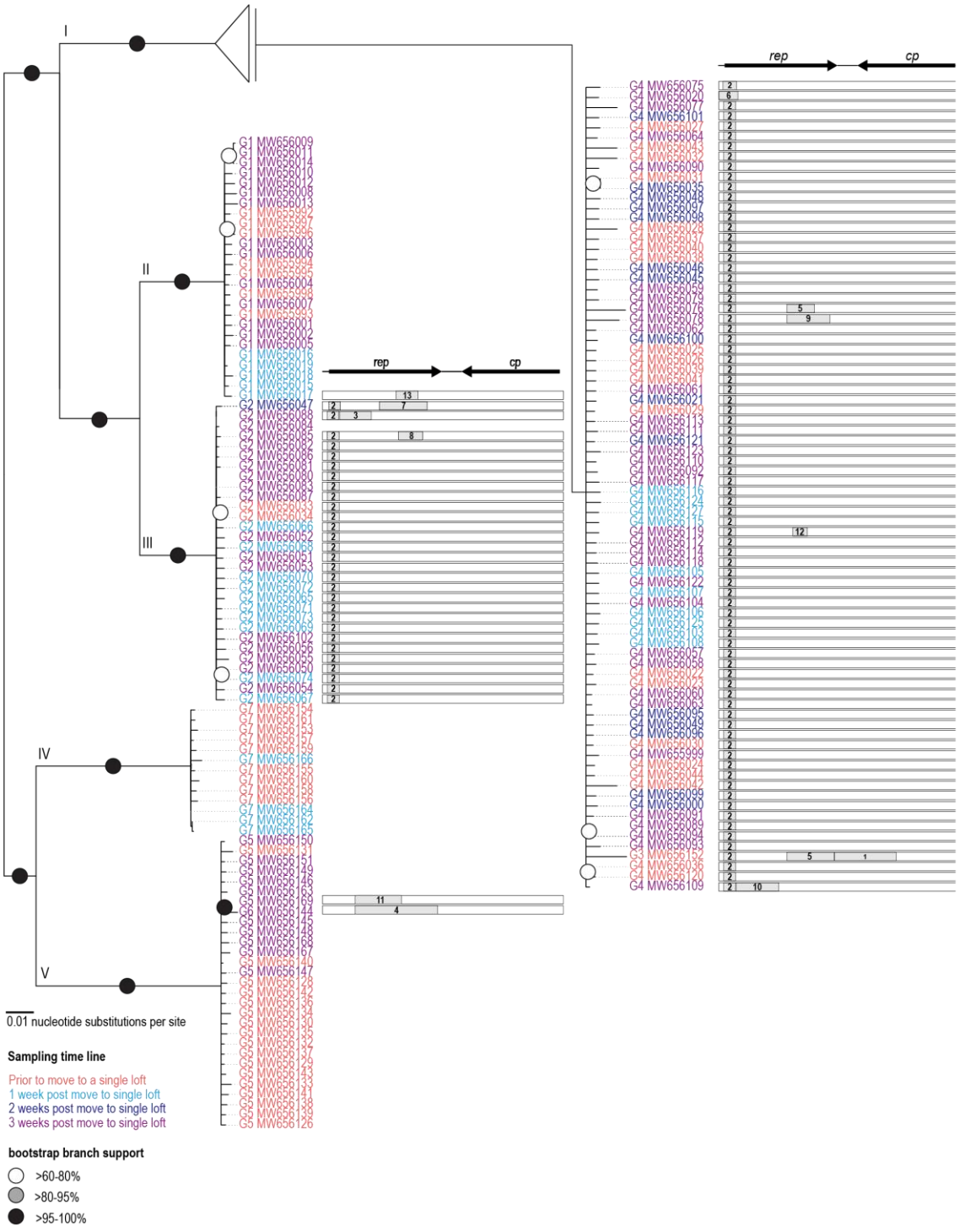


Figure 3.4: Maximum Likelihood phylogenetic tree of PiCV sequences recovered from study with recombinant regions removed. Recombinant regions identified in the genomes are show next to the accession numbers. The five main clades labelled in roman numerals (I-V). Genotypes (G1-G7) are labelled next to all the accession numbers.

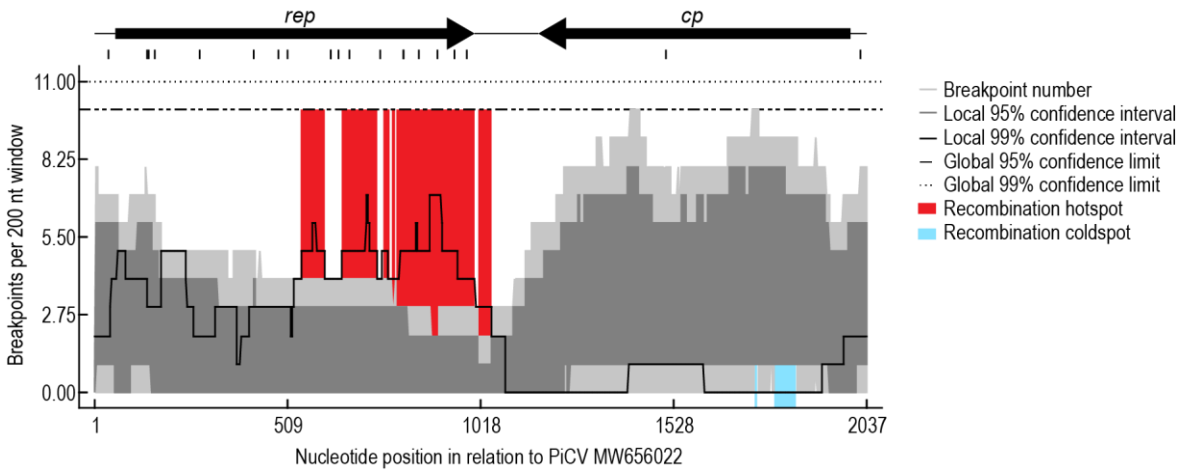


Figure 3.5: Recombination break point analysis for PiCV genomes from this study. Hot-spots and cold-spots are highlighted. Genome organization based on PiCV sequence MW656022 as a representative.

Conclusion

In this study we simulate the viral dynamics at play when infected and non-infected animals are acquired from different sources and then house together. We show that genotypes of PiCV to non-infected birds and also to already infected birds either resulting in co-infection or displacement of previous genotype, e.g. the genotype 4 in TSP7 appears to have "out competed" genotypes 1, 2 and 7 (Figure 3.2). We also demonstrate that recombination plays a role in virus evolution, although appears to be relatively smaller with different recombination hots spots compared that observed in global dataset with sequences with a larger temporal signal of decades. Nonetheless, this study demonstrates the PiCV spreads in animals within a loft within a short time frame, the animals can harbor multiple genotypes and dominant genotypes possibly outcompete previous variants.

CHAPTER 4

NOVEL VIRUSES BELONGING TO THE CIRCOVIRIDAE FAMILY RECOVERED FROM WILD WATERFOWL

Abstract

Kiwanis Park is a man-made lake located in Tempe Arizona which is home to multiple avian and fish species. Using a metagenomic approach, two novel circoviruses were identified in fecal samples from waterfowl collected at Kiwanas Park. One novel circovirus tentatively named Anatidae associated circovirus 1 (AnaACV1) was recovered from eight different fecal samples and the other, referred to as Anatidae associated circovirus 2 (AnaACV2) was recovered from two different samples. AnaACV1 shares 60-64% full genome similarity to duck circovirus, goose circovirus, and swan circovirus. AnaACV2 shares 71% full genome similarity to bat associated circovirus 3 and 4. This adds to the global circovirus dataset and allows for a better understanding of the viruses circulating within the local ecological communities.

Introduction

Due to advancements in metagenomic sequencing, namely the coupling of rolling-circle amplification and high-throughput, there has been an increase in the number of novel circular single-stranded DNA viruses discovered. A large portion of these viruses belong to the newly established phylum *Cressdnaviricota* (Krupovic et al., 2020). This phylum is composed of seven families; *Bacilladnaviridae, Circoviridae, Geminiviridae, Genomoviridae, Nanoviridae, Redondoviridae, and Smacoviridae* (Krupovic et al., 2020).

Circoviruses, (family *Circovirus*) have a circular, single-stranded DNA (ssDNA) of 1.7-2.1kb that encode a viral capsid protein (CP), replication-associated protein (Rep),

and an origin of replication (*ori*). Circoviruses have been shown to infect pigs, parrots, goose, ducks, and pigeons (Ball et al., 2004; Hattermann et al., 2003; Pass & Perry, 1984; Segalés, 2012; Leslie W Woods et al., 1994). Circovirus infection in ducks and geese can lead to several negative health outcomes, including stunted growth and feather disorders (Hattermann et al., 2003; D Todd et al., 2001). Waterfowl, such as ducks and geese, have been extensively surveyed for several viruses, with a focus on influenza (Halvorson et al., 1983; Markwell & Shortridge, 1982; Wallensten et al., 2007). Freshwater bodies of water are areas of high biodiversity that are important for conservation of many species that rely on them (Dudgeon et al., 2006; Strayer & Dudgeon, 2010). This lake is inhabited by multiple avian species and fish species. One that is highly prevent is the American wigeon (*Mareca americana*), belonging to the family *Anatidae*, they are a species of duck found with widespread distribution across North America, including Kiwanis Park. Previous studies have described avian influenza virus in American wigeons (Runstadler et al., 2007; Siembieda et al., 2010), but no other viruses associated with American wigeon have been documented (as of March 31, 2021).

In this study we used a metagenomic approach to identify ten virus genomes that represent two novel circovirus species from fecal samples of American wigeon collected at a man-made lake in Kiwanis park located in Tempe, Arizona, USA.

Materials and Methods

A sample (K19) of American wigeon feces was collected on 13th Jan 2021 as part of a preliminary analysis of viruses associated with American wigeon. Thereafter, more samples of American wigeon W1-19 were collected on 1st March 2021. Viral DNA was extracted from American wigeon feces using High Pure viral nucleic acid kit (Roche

Diagnostics, USA) according to manufacturer's instructions. Viral DNA was amplified using rolling circular amplification (RCA) with TempliPhi 2000 kit (GE Healthcare, USA). The resulting RCA of sample K19 was used to prepare Illumina sequencing libraries which were sequenced on the Illumina 4000 platform at Macrogen Inc. (Korea). The raw reads were *de nov*o assembled using SPAdes v 3.12.0 (Anton Bankevich et al., 2012) and contigs (>750nts) were analyzed using BLASTx (Altschul, Gish, Miller, Myers, & Lipman, 1990) against a viral protein database. Contigs were sorted based upon similarities to viral families. Abutting primers (AnaACV1_F: 5'-CACCCTCACCCATCTGATTCTATAATAGTC-3' R:5'-GAGATATGGCCTACCTTTCAGATTTTC-3'. AnaACV2_F: 5'-GAGTGGTATGACA-AAGAGAAGATACCC-3' R: 5'-CTCTGGCCTCTTATTGGACGTTATAATTAC-3'.) were designed based on these circular ssDNA virus-like contigs. Primers were used to screen and recover the viral sequences from samples W1-19 using PCR with Kapa Hifi Hotstart DNA polymerase (Kapa Biosystems, USA) following the manufacturer recommended thermal cycling conditions with an annealing temperature of 55 °C and 60 °C, respectively. Amplified genomes were resolved in 0.7% agarose gel, excised and purified. The purified amplicons were ligated with pJET 1.2 vector (Thermo Fisher Scientific, USA) and transformed into *Escerichia coli* DH5-alpha competent cells. The purified recombinant plasmids were Sanger sequenced at Macrogen Inc. (Korea) by primer walking. The contigs were assembled and annotated using Geneious v11.0.3 (Biomatters Ltd. New Zealand).

Host identification was carried out using PCR amplification of the cytochrome C oxidase 1 (COI) gene with primers described in (Hebert, Stoeckle, Zemlak, & Francis,

2004). PCR amplification was performed with Kapa Blood PCR Mix B using thermal cycling conditions of 95 °C for 5 minutes followed by 5 cycles of 30 seconds at 95 °C, 30 seconds at 62 °C, and 40 seconds at 72 °C, followed by 5 cycles of 30 seconds at 95 °C, 30 seconds at 66 °C, and 40 seconds at 72 °C, followed by 20 cycles of 30 seconds at 95 °C, 30 seconds at 52 °C, and 40 seconds at 72 °C.

Sequences were aligned with MAFFT (Kazutaka Katoh & Daron M Standley, 2013) and this alignment was used to infer a maximum likelihood phylogenetic tree using PhyML 3.0 (Guindon et al., 2009) with GTR+I+G nucleotide substitution model for full genome analysis determine using jModelTest (Diego Darriba, Taboada, Doallo, & Posada, 2012). The genomes for Rep and CP protein sequences of these were aligned using MAFFT (Kazutaka Katoh & Daron M Standley, 2013) and maximum likelihood phylogenetic tree using PhyML 3.0 with LG+I+G (for Rep alignment) and VT+I+G+F (for CP alignment) substitution models determined as best fit models using ProtTest (D Darriba, Taboada, Doallo, & Posada, 2011). The trees were rooted using cyclovirus sequences as an outgroup. Branch support <0.8 aLRT support were collapsed using TreeGraph2 (Stöver & Müller, 2010)

All genome-wide and protein specific pairwise identities were determined using SDT 1.2 (Muhire et al., 2014).

Results and Discussion

Wigeon fecal samples were visually identified during collection and PCR of the cytochrome C oxidase 1 (COI) gene used to confirm host ID (Hebert et al., 2004).

A total of 10 circovirus genomes were recovered, all of which encode for a CP, a nonanucleotide motif "TA(T/G)TATTAC" origin of replication, and a Rep protein that contains conserved motifs essential for rolling-circle replication, RCR motifs, (Ilyina & Koonin, 1992) and DNA binding and cleavage, HUH endonuclease motifs (Chandler et al., 2013) (Table 4.1, Figure 4.1). These samples were then screened for two novel circoviruses using back-to-back primers. Eight genomes were recovered from eight samples (K19, W1, W4, W6, W9, W13, W15, W18), and these shared > 99% full genome pairwise identity. Further these genomes share 64% full genome pairwise identity with duck circovirus (DuCV DQ100076) (Banda et al., 2007), 63% Rep amino acid identity with DuCV (DQ100076) and goose circovirus (GoCV AJ304456) (D Todd et al., 2001), and a 39% CP amino acid identity with bat associated circovirus 4 (BatACV4 KT783484) (Lima et al., 2015). Based on an 80% genome-wide sequence identity demarcation outlined in Rosario et al. (2017), these genomes belong to a new species tentatively named Anatidae associated circovirus 1 (AnaACV1). Phylogenetic analysis of AnaACV1 full genome compared to a representative dataset current circovirus species show AnaACV1 forms a clade that are sister taxa with the clade formed by duck circovirus (DQ100076), goose circovirus (AJ304456), and swan circovirus (SwCV EU056309) (Halami et al., 2008) (Figure 4.1). Phylogeny of the Rep sequences shows AnaACV1 forming a clade that sits most closely to those containing the DuCV, GoCV, and SwCV. Lastly, the CP phylogeny shows AnaACV1 being sister taxa with DuCV, GoCV, and SwCV clade and bat associated circovirus 3 (BatACV3;JQ814849) (Wu et al., 2012) and bat associated circovirus 4 (BatACV4; KT783484) recovered from the greater horseshoe bat, *Rhinolophus ferrumequinum*, and the Mexican free-tailed bat, *Tadarida brasiliensis*, respectively.

Two unique circovirus genomes (AnaACV2: K19, W3) sharing >97% full genome identity were recovered from two American wigeon fecal samples. AnaACV2 shares 71% full genome pairwise identity with BatACV3 (JQ814849) and BatACV4 (KT783484). AnaACV2 shares an 82% Rep amino acid identity with BatACV3 and a 64% CP amino acid identity with BatACV3 and BatACV4. These genomes tentatively belong to a new species tentatively named Anatidae associated circovirus 2 (AnaACV2). Phylogenetic analysis of AnaACV2 full genome, Rep protein, and CP show AnaCV2 sits in a clade with BatACV3 and BatACV4.

AnaACV1 and AnaCV2 share 57% full genome identity, 48% Rep protein amino acid identity, and 34-35% CP amino acid identity. One sample, K19 which was collected in the initial screening process harbored both AnaACV1 and 2.

Virus	Motif 1	Motif 2	Motif 3	Walker A	Walker B	Motic C	Arg Finger	Nonanucleotide
AnaACV1	VFTINN	PHLQ	DNKAYCQK	GPPGCGKSRWAW	VMDDF	VTSN	ALYRRI	TATTATTAC
AnaACV2	CFTINN	KHLQ	QNKIYCSK	GEPGSGKSRFAN	ILDDF	ITSN	AMWRRF	TAGTATTAC

Table 4.1: Summary of viruses recovered from the study and HUH endonucleases and superfamily 3 helicases identified in the replication-associated proteins.

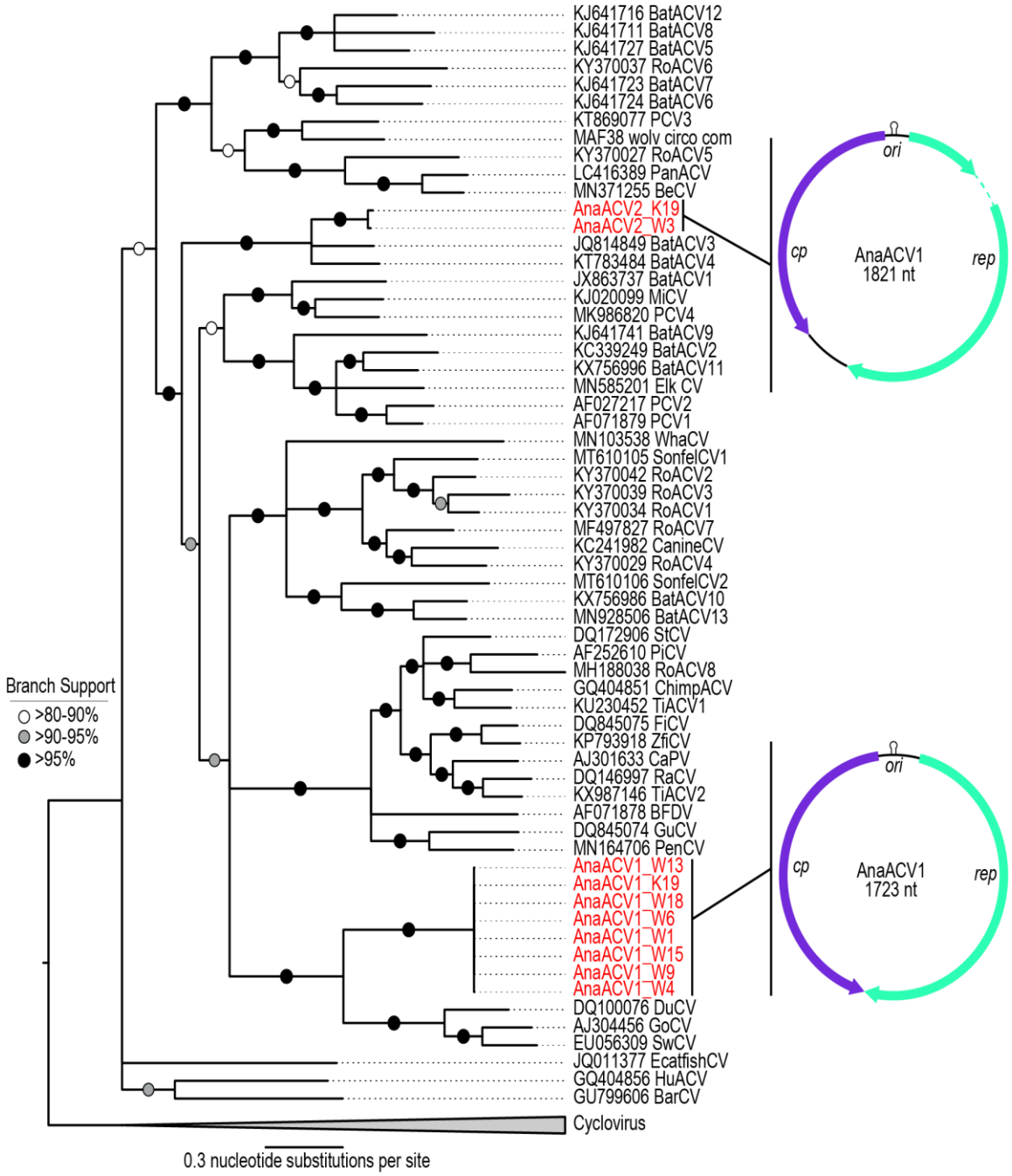


Figure 4.1: Phylogenetic tree of the aligned genome of a representative of all circovirus available in GenBank. AnaACV genomes highlighted in red. Cartoon depiction of circoviruses genome organization shown.

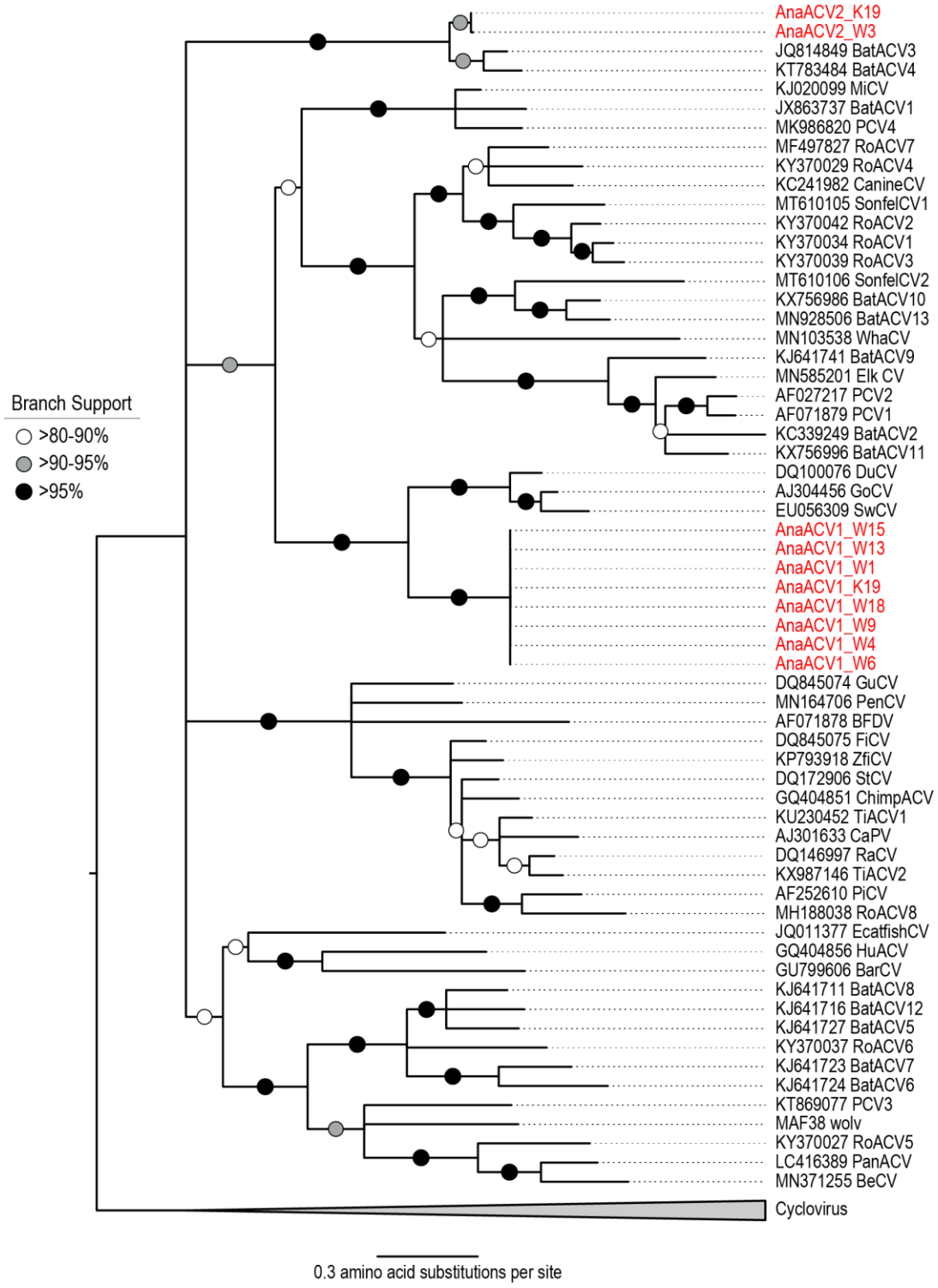


Figure 4.2: Phylogenetic tree of the aligned Rep protein encoded by a representative of all circoviruses available in GenBank. AnaACV genomes highlighted in red.

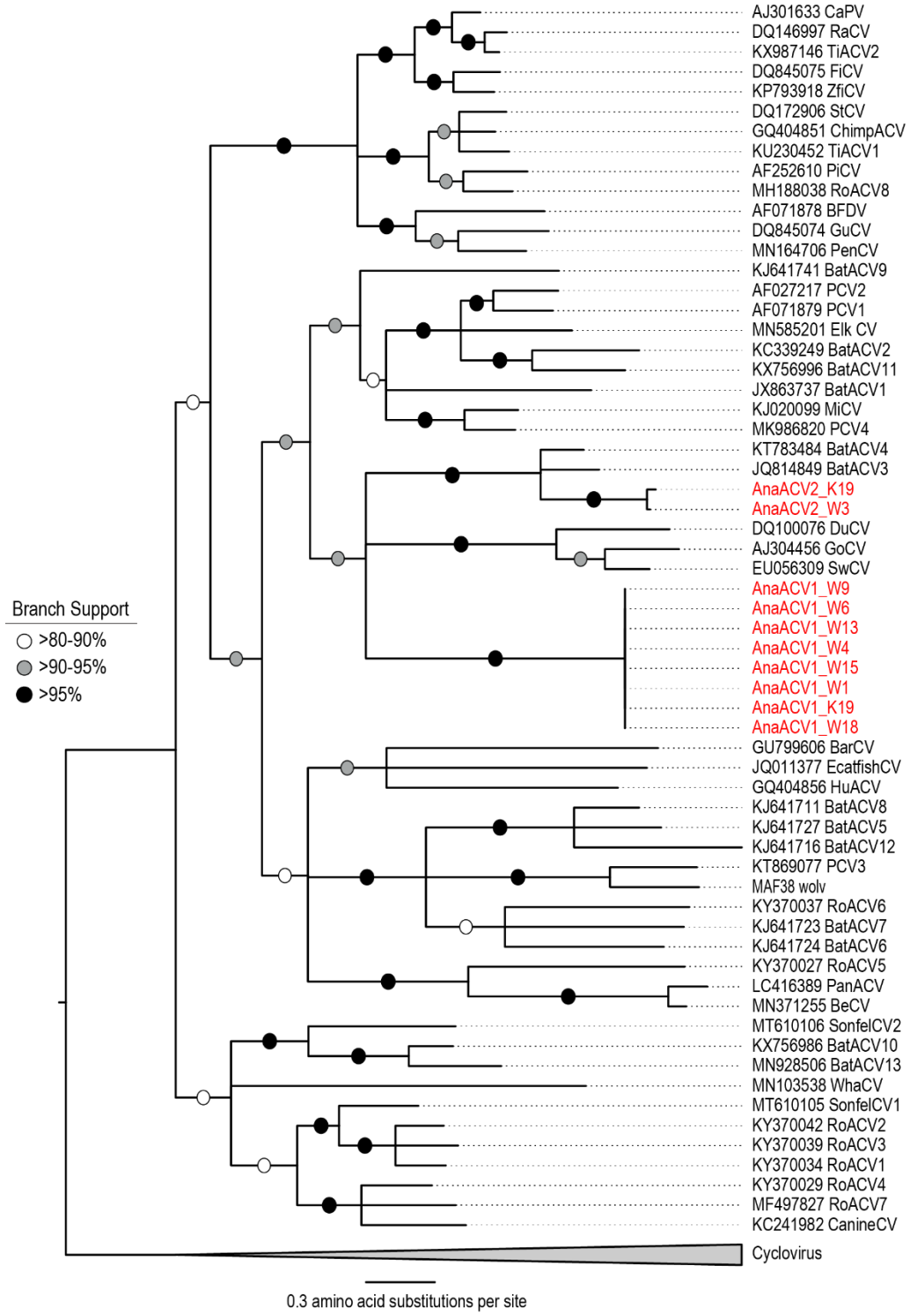


Figure 4.3: Phylogenetic tree of the aligned CP protein encoded by a representative of all circovirus available in GenBank. AnaACV genomes highlighted in red.

Conclusion

The aim of this study was to explore the circoviruses that are circulating in waterfowl living in Tempe, Arizona. Given circoviruses are known to infect multiple avian species, including waterfowl, it is highly likely these novel circoviruses infect American wigeons, however, this needs to be further studied. A total of 10 genomes were recovered falling into two tentative species, AnaACV1 sharing relatedness to duck circovirus, goose circovirus, and swan circovirus and AnaACV2 sharing relatedness to bat associated circovirus 3 and bat associated circovirus 4. These two novel species of circoviruses identified in American wigeon increases our knowledge of circovirus diversity and potential host range.

CHAPTER 5

CONCLUSION

The aim of the study described in Chapter 2 was to identify circular DNA viruses associated with pigeons primarily in cloacal swab samples. Using metagenomic approaches, a genome that shared relatedness to viruses belonging to viruses in the *Phapecoctavirus* genus was identified. This virus was named domestic pigeon swab associated virus (dompiswa virus). Dompiswa virus is 150,892 bps in length and shares 92.3% inter-genomic distance with *Klebsiella* phage ZCKP1. Five of the 16 viruses belonging to the *Phapecoctavirus* genus have been shown to infect enterobacteria *Campylobacter jejuni, Escherichia coli* or *Klebsiella pneumoniae,* therefore it is likely that dompiswa virus infects an enterobacteria, but further studies are needed to confirm the host.

The aim of the study described in Chapter 3 was to determine the viral dynamics of pigeon circovirus with infected and noninfected animals, from different origins. It is shown that PiCV genotypes can infect uninfected animals and can infect already infected animals, resulting in co-infection in some animals and competition causing one genotype to out compete another. Recombination patterns within the study did not match the previous studies that determined hotspots in a global dataset of PiCV, as recombination seemed to play a smaller role than previously observed. PiCV was still observed to spread within the population to uninfected animals within a short time span, as well as showing animals can harbor multiple genotypes as well as competition amongst genotypes being observed.

The aim of the research described in Chapter 4 was to screen waterfowl samples from Kiwanis Park in Tempe Arizona for novel circoviruses. A total of 10 genomes were

recovered, representing two novel circovirus species, named AnaACV1 and AnaACV2. AnaACV1 shares relatedness to duck circovirus, goose circovirus, and swan circovirus of 60 – 64%. Whereas AnaACV2 shares relatedness to bat associated circovirus 3 and bat associated circovirus 4 with 71%. Previous studied have determined circovirus infection amongst other species of waterfowl, therefore it is likely that these viruses infect American wigeon, however these claims need to be studied. Nonetheless, these associated viruses increase our knowledge of circovirus diversity and hot range.

The overall objective of the research described in this thesis was to gain a deeper understanding of the diversity and evolution of circular DNA viruses associated with avian species. This was achieved by analysis of novel DNA viruses and analysis of recombination of a known virus. Further directions of this work involve identifying host species for dompiswa virus, as well as confirming host for AnaACV1 and AnaACV2. For PiCV, expanding on recombination studies over a short period of time, a long-term recombination study could be informative to investigate how the distribution of genotypes changes and to understand the driving force behind that change. Also, looking at a large set of environmental samples to understand recombination and selection dynamics in a noncaptive setting.

REFERENCES

Chapter 1:

A. Stenzel, T., Pestka, D., Tykałowski, B., Śmiałek, M., & Koncicki, A. (2012). Epidemiological investigation of selected pigeon viral infections in Poland. *Veterinary Record, 171*(22), 562-562.

Amery-Gale, J., Marenda, M. S., Owens, J., Eden, P. A., Browning, G. F., & Devlin, J. M. (2017). A high prevalence of beak and feather disease virus in non-psittacine Australian birds.

Banda, A., Galloway-Haskins, R. I., Sandhu, T. S., & Schat, K. A. (2007). Genetic analysis of a duck circovirus detected in commercial Pekin ducks in New York. *Avian diseases, 51*(1), 90-95.

Bartlett, J. M., & Stirling, D. (2003). A short history of the polymerase chain reaction *PCR protocols* (pp. 3-6): Springer.

Beach, N. M., & Meng, X.-J. (2012). Efficacy and future prospects of commercially available and experimental vaccines against porcine circovirus type 2 (PCV2). *Virus research, 164*(1-2), 33-42.

Bonne, N., Shearer, P., Sharp, M., Clark, P., & Raidal, S. (2009). Assessment of recombinant beak and feather disease virus capsid protein as a vaccine for psittacine beak and feather disease. *Journal of General Virology, 90*(3), 640-647.

Chandler, M., De La Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., & Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews Microbiology, 11*(8), 525-538.

Chen, C.-L., Chang, P.-C., Lee, M.-S., Shien, J.-H., Ou, S.-J., & Shieh, H. (2003). Nucleotide sequences of goose circovirus isolated in Taiwan. *Avian Pathology, 32*(2), 165-171.

Chen, C.-L., Wang, P.-X., Lee, M.-S., Shien, J.-H., Shieh, H. K., Ou, S.-J., . . . Chang, P.-C. (2006). Development of a polymerase chain reaction procedure for detection and differentiation of duck and goose circovirus. *Avian diseases, 50*(1), 92-95.

Coletti, M., Franciosini, M., Asdrubali, G., & Passamonti, F. (2000). Atrophy of the primary lymphoid organs of meat pigeons in Italy associated with circoviruslike particles in the bursa of Fabricius. *Avian diseases*, 454-459.

Crowther, R., Berriman, J., Curran, W., Allan, G., & Todd, D. (2003). Comparison of the structures of three circoviruses: chicken anemia virus, porcine circovirus type 2, and beak and feather disease virus. *Journal of virology, 77*(24), 13036-13041.

Das, S., Smith, K., Sarker, S., Peters, A., Adriaanse, K., Eden, P., . . . Raidal, S. R. (2019). Assessing circovirus gene flow in multiple spill-over events. *Virus genes, 55*(6), 802-814.

Duchatel, J.-P., Todd, D., Curry, A., Smyth, J., Bustin, J., & Vindevogel, H. (2005). New data on the transmission of pigeon circovirus. *The Veterinary Record, 157*(14), 413.

Duchatel, J.-P., Todd, D., Smyth, J., Bustin, J., & Vindevogel, H. (2006). Observations on detection, excretion and transmission of pigeon circovirus in adult, young and embryonic pigeons. *Avian Pathology, 35*(1), 30-34.

Duchatel, J.-P., Todd, D., Willeman, C., & Losson, B. (2009). Quantification of pigeon circovirus in serum, blood, semen and different tissues of naturally infected pigeons using a real-time polymerase chain reaction. *Avian Pathology, 38*(2), 143-148.

Franciosini, M., Fringuelli, E., Tarhuni, O., Guelfi, G., Todd, D., Proietti, P. C., . . . Asdrubali, G. (2005). Development of a polymerase chain reaction-based in vivo method in the diagnosis of subclinical pigeon circovirus infection. *Avian diseases, 49*(3), 340-343.

Gillespie, J., Opriessnig, T., Meng, X., Pelzer, K., & Buechner-Maxwell, V. (2009). Porcine circovirus type 2 and porcine circovirus-associated disease. *Journal of Veterinary Internal Medicine, 23*(6), 1151-1163.

Guo, J., Tian, J., Tan, X., Yu, H., Ding, S., Sun, H., & Yu, X. (2011). Pathological observations of an experimental infection of geese with goose circovirus. *Avian Pathology, 40*(1), 55-61.

Gutierrez, C. (1999). Geminivirus DNA replication. *Cellular and Molecular Life Sciences CMLS, 56*(3), 313-329.

Haible, D., Kober, S., & Jeske, H. (2006). Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of virological methods, 135*(1), 9-16.

Halami, M., Nieper, H., Müller, H., & Johne, R. (2008). Detection of a novel circovirus in mute swans (Cygnus olor) by using nested broad-spectrum PCR. *Virus research, 132*(1-2), 208-212.

Hattermann, K., Schmitt, C., Soike, D., & Mankertz, A. (2003). Cloning and sequencing of Duck circovirus (DuCV). *Archives of Virology, 148*(12), 2471-2480.

Heath, L., Martin, D. P., Warburton, L., Perrin, M., Horsfield, W., Kingsley, C., . . . Williamson, A.-L. (2004). Evidence of unique genotypes of beak and feather disease virus in southern Africa. *Journal of virology, 78*(17), 9277-9284.

Huang, J., Yang, C., Jia, R., Wang, M., Chen, S., Liu, M., . . . Wu, Y. (2018). Induction of a protective response in ducks vaccinated with a DNA vaccine encoding engineered duck circovirus Capsid protein. *Veterinary microbiology, 225*, 40-47.

Huang, Y.-L., Castaneda, O. A., Thongchan, D., Khatri-Chhetri, R., Tsai, S.-S., & Wu, H.-Y. (2017). Pigeon circovirus infection in disqualified racing pigeons from Taiwan. *Avian Pathology, 46*(4), 359-366.

Ilyina, T. V., & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic acids research, 20*(13), 3279-3285.

Johne, R., Fernandez-de-Luco, D., Höfle, U., & Müller, H. (2006). Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification. *Journal of General Virology, 87*(5), 1189-1195.

Kazlauskas, D., Varsani, A., Koonin, E. V., & Krupovic, M. (2019). Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nature communications, 10*(1), 1-12.

Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., . . . Kuhn, J. H. (2020). Global organization and proposed megataxonomy of the virus world. *Microbiology and molecular biology reviews, 84*(2).

Krupovic, M., Varsani, A., Kazlauskas, D., Breitbart, M., Delwart, E., Rosario, K., . . . Zerbini, F. M. (2020). Cressdnaviricota: a virus phylum unifying 7 families of Rep-encoding viruses with single-stranded, circular DNA genomes. *Journal of virology*.

Ledwoń, A., Bailey, T., O'DONOVAN, D., Mckeown, S., Lloyd, C., WIĘCKOWSKI, T., . . . Wernery, U. (2011). Prevalence of circovirus and adenovirus in pigeons in Dubai. *Med Weter, 67*, 752-756.

Loiko, M., Junqueira, D., Varela, A., Tochetto, C., Scheffer, C., Lima, D., . . . Mayer, F. Q. (2018). Columbid circoviruses detected in free ranging pigeons from Southern Brazil: insights on PiCV.

Ma, C.-M., Hon, C.-C., Lam, T.-Y., Li, V. Y.-Y., Wong, C. K.-W., de Oliveira, T., & Leung, F. C.-C. (2007). Evidence for recombination in natural populations of porcine circovirus type 2 in Hong Kong and mainland China. *Journal of General Virology, 88*(6), 1733-1737.

Martin, D. P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., & Varsani, A. (2011). Recombination in eukaryotic single stranded DNA viruses. *Viruses, 3*(9), 1699-1738.

Mohsen, M. G., & Kool, E. T. (2016). The discovery of rolling circle amplification and rolling circle transcription. *Accounts of chemical research, 49*(11), 2540-2550.

Nath, B. K., Das, S., Roby, J. A., Sarker, S., Luque, D., Raidal, S. R., & Forwood, J. K. (2021). Structural Perspectives of Beak and Feather Disease Virus and Porcine Circovirus Proteins. *Viral Immunology, 34*(1), 49-59.

Niu, X., Liu, L., Han, C., Li, J., & Zeng, X. (2018). First findings of duck circovirus in migrating wild ducks in China. *Veterinary microbiology, 216*, 67-71.

Opriessnig, T., Meng, X.-J., & Halbur, P. G. (2007). Porcine circovirus type 2–associated disease: update on current terminology, clinical manifestations, pathogenesis, diagnosis, and intervention strategies. *Journal of Veterinary Diagnostic Investigation, 19*(6), 591-615.

Pass, D., & Perry, R. (1984). The pathology of psittacine beak and feather disease. *Australian Veterinary Journal, 61*(3), 69-74.

Phenix, K. V., Weston, J. H., Ypelaar, I., Lavazza, A., Smyth, J. A., Todd, D., . . . Raidal, S. R. (2001). Nucleotide sequence analysis of a novel circovirus of canaries and its relationship to other members of the genus Circovirus of the family Circoviridae. *Journal of General Virology, 82*(11), 2805-2809.

Raidal, S., & Cross, G. (1994). Control by vaccination of psittacine beak and feather disease in a mixed flock of Agapornis spp. *Australian Veterinary Practitioner, 24*(4), 178-180.

Raidal, S., Sarker, S., & Peters, A. (2015). Review of psittacine beak and feather disease and its effect on Australian endangered species. *Australian Veterinary Journal, 93*(12), 466-470.

Raue, R., Schmidt, V., Freick, M., Reinhardt, B., Johne, R., Kamphausen, L., . . . Krautwald-Junghanns, M.-E. (2005). A disease complex associated with pigeon circovirus infection, young pigeon disease syndrome. *Avian Pathology, 34*(5), 418-425.

Reitz, N., Kamphausen, L., Krautwald-Junghanns, M., Makert, C., & Schmidt, V. (2003). Fragebogenaktion Jungtauben-Krankheit. *Die Brieftaube, 35*, 1291-1294.

Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-throughput sequencing technologies. *Molecular cell, 58*(4), 586-597.

Rinder, M., Schmitz, A., Peschel, A., & Korbel, R. (2015). Complete genome sequence of a novel circovirus from zebra finch. *Genome announcements, 3*(3).

Rosario, K., Breitbart, M., Harrach, B., Segalés, J., Delwart, E., Biagini, P., & Varsani, A. (2017). Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Archives of Virology, 162*(5), 1447-1463.

Rosario, K., Marinov, M., Stainton, D., Kraberger, S., Wiltshire, E. J., Collings, D. A., . . . Varsani, A. (2011). Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). *Journal of General Virology, 92*(6), 1302-1308.

Sarker, S., Das, S., Ghorashi, S. A., Forwood, J. K., & Raidal, S. R. (2019). Pigeon circoviruses from feral pigeons in Australia demonstrate extensive recombination and genetic admixture with other circoviruses. *Avian Pathology, 48*(6), 512-520.

Sarker, S., Ghorashi, S. A., Swarbrick, C. M., Khandokar, Y. B., Himiari, Z., Forwood, J. K., & Raidal, S. R. (2015). An efficient approach for recombinant expression and purification of the viral capsid protein from beak and feather disease virus (BFDV) in Escherichia coli. *Journal of virological methods, 215*, 1-8.

Sarker, S., Lloyd, C., Forwood, J., & Raidal, S. R. (2016). Forensic genetic evidence of beak and feather disease virus infection in a Powerful Owl, Ninox strenua. *Emu, 116*(1), 71-74.

Sarker, S., Moylan, K. G., Ghorashi, S. A., Forwood, J. K., Peters, A., & Raidal, S. R. (2015). Evidence of a deep viral host switch event with beak and feather disease virus infection in rainbow bee-eaters (Merops ornatus). *Scientific Reports, 5*(1), 1-8.

Schoemaker, N., Dorrestein, G., Latimer, K., Lumeij, J., Kik, M., Van Der Hage, M., & Campagnoli, R. (2000). Severe leukopenia and liver necrosis in young African grey parrots (Psittacus erithacus erithacus) infected with psittacine circovirus. *Avian diseases*, 470-478.

Segalés, J. (2012). Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis. *Virus research, 164*(1-2), 10-19.

Smits, S. L., Zijlstra, E., van Hellemond, J. J., Schapendonk, C. M., Bodewes, R., Schürch, A. C., . . . Osterhaus, A. D. (2013). Novel cyclovirus in human cerebrospinal fluid, Malawi, 2010–2011. *Emerging infectious diseases, 19*(9), 1511.

Soike, D., Kohler, B., & Albrecht, K. (1999). A circovirus-like infection in geese related to a runting syndrome. *Avian Pathology, 28*(2), 199-202.

Stenzel, T., Dziewulska, D., Muhire, B. M., Hartnady, P., Kraberger, S., Martin, D. P., & Varsani, A. (2018). Recombinant goose circoviruses circulating in domesticated and wild geese in Poland. *Viruses, 10*(3), 107.

Stenzel, T., Dziewulska, D., Tykałowski, B., Śmiałek, M., Kowalczyk, J., & Koncicki, A. (2018). Immunogenicity of pigeon circovirus recombinant capsid protein in pigeons. *Viruses, 10*(11), 596.

Stenzel, T., Piasecki, T., Chrząstek, K., Julian, L., Muhire, B. M., Golden, M., . . . Varsani, A. (2014). Pigeon circoviruses display patterns of recombination, genomic secondary structure and selection similar to those of beak and feather disease viruses. *Journal of General Virology, 95*(6), 1338-1351.

Stewart, M. E., Perry, R., & Raidal, S. R. (2006). Identification of a novel circovirus in Australian ravens (Corvus coronoides) with feather disease. *Avian Pathology, 35*(02), 86-92.

Todd, D. (2004). Avian circovirus diseases: lessons for the study of PMWS. *Veterinary microbiology, 98*(2), 169-174.

Todd, D., Weston, J., Soike, D., & Smyth, J. (2001). Genome sequence determinations and analyses of novel circoviruses from goose and pigeon. *Virology, 286*(2), 354-362.

Tsai, S. S., Chang, Y. L., Huang, Y. L., Liu, H. J., Ke, G. M., Chiou, C. J., . . . Chuang, K. P. (2014). Development of a loop-mediated isothermal amplification method for rapid detection of pigeon circovirus. *Archives of Virology, 159*(5), 921-926.

Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Dempsey, D. M., . . . Hendrickson, R. C. (2020). Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020): Springer.

Wittwer, C. T., Herrmann, M. G., Moss, A. A., & Rasmussen, R. P. (1997). Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques, 22*(1), 130-138.

Woods, L. W., Latimer, K. S., Niagro, F. D., Riddell, C., Crowley, A. M., Anderson, M. L., . . . Nordhausen, R. W. (1994). A retrospective study of circovirus infection in pigeons: nine cases (1986–1993). *Journal of Veterinary Diagnostic Investigation, 6*(2), 156-164.

Yu, X., Zhu, C., Zheng, X., He, S., & Liu, X. (2007). Genome analysis and epidemiological investigation of goose circovirus detected in eastern China. *Virus genes, 35*(3), 605-609.

Zhaolong, L., Guanghua, F., Zhihua, F., Jianhua, C., Shaohua, S., Rongchang, L., . . . Yu, H. (2020). Evaluation of a novel inactivated vaccine against duck circovirus in muscovy ducks. *Veterinary microbiology*, *241*, 108574.

Zhou, X., Ren, L., Li, Y., Zhang, M., Yu, Y., & Yu, J. (2010). The next-generation sequencing technology: a technology review and future perspective. *Science China Life Sciences, 53*(1), 44-57.

Chapter 2:

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol, 19*(5), 455-477. doi:10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., . . . Xia, F. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep, 5*, 8365. doi:10.1038/srep08365

Brister, J. R., Ako-Adjei, D., Bao, Y., & Blinkova, O. (2015). NCBI viral genomes resource. *Nucleic Acids Res, 43*(Database issue), D571-577. doi:10.1093/nar/gku1207

Guindon, S., Delsuc, F., Dufayard, J.-F., & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML *Bioinformatics for DNA sequence analysis* (pp. 113-137): Springer.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Lavigne, R., & Ceyssens, P.-J. (2012). Myoviridae. In A. M. Q. King, M. J. Adams, E. B. Carstens, & E. J. Lefkowitz (Eds.), *Virus Taxonomy* (pp. 46-62). San Diego: Elsevier.

Moraru, C., Varsani, A., & Kropinski, A. M. (2020). VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses, 12*(11), 1268. doi:10.3390/v12111268

Posada, D., & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics, 14*(9), 817-818. doi:10.1093/bioinformatics/14.9.817

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ, 3*, e985. doi:10.7717/peerj.985

Stover, B. C., & Muller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics, 11*(1), 7. doi:10.1186/1471-2105-11-7

Taha, O. A., Connerton, P. L., Connerton, I. F., & El-Shibiny, A. (2018). Bacteriophage ZCKP1: A Potential Treatment for Klebsiella pneumoniae Isolated From Diabetic Foot Patients. *Front Microbiol, 9*, 2127. doi:10.3389/fmicb.2018.02127

Chapter 3:

Abadie, J., Nguyen, F., Groizeleau, C., Amenna, N., Fernandez, B., Guereaud, C., . . . Wyers, M. (2001). Pigeon circovirus infection: pathological observations and suggested pathogenesis. *Avian Pathol, 30*(2), 149-158. doi:10.1080/03079450124811

Agnihotri, K., Smith, C., Oakey, J., & Storie, G. (2021). Pigeon adenovirus and pigeon torque teno virus associated with acute multifocal hepatic necrosis in pigeons in Queensland, Australia. *Arch Virol*. doi:10.1007/s00705-021-05033-x

Coletti, M., Franciosini, M. P., Asdrubali, G., & Passamonti, F. (2000). Atrophy of the primary lymphoid organs of meat pigeons in Italy associated with circoviruslike particles in the bursa of Fabricius. *Avian Dis, 44*(2), 454-459.

Duchatel, J. P., Todd, D., Curry, A., Smyth, J. A., Bustin, J. C., & Vindevogel, H. (2005). New data on the transmission of pigeon circovirus. *Vet Rec, 157*(14), 413-415. doi:10.1136/vr.157.14.413

Duchatel, J. P., Todd, D., Smyth, J. A., Bustin, J. C., & Vindevogel, H. (2006). Observations on detection, excretion and transmission of pigeon circovirus in adult, young and embryonic pigeons. *Avian Pathol, 35*(1), 30-34. doi:10.1080/03079450500465692

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Franciosini, M. P., Fringuelli, E., Tarhuni, O., Guelfi, G., Todd, D., Casagrande Proietti, P., . . . Asdrubali, G. (2005). Development of a polymerase chain reaction-based in vivo method in the diagnosis of subclinical pigeon circovirus infection. *Avian Dis, 49*(3), 340-343. doi:10.1637/7334-012705R.1

Freick, M., Muller, H., & Raue, R. (2008). Rapid detection of pigeon herpesvirus, fowl adenovirus and pigeon circovirus in young racing pigeons by multiplex PCR. *J Virol Methods, 148*(1-2), 226-231. doi:10.1016/j.jviromet.2007.11.003

Haag-Wackernagel, D., & Moch, H. (2004). Health hazards posed by feral pigeons. *J Infect, 48*(4), 307-313. doi:10.1016/j.jinf.2003.11.001

Huang, Y. L., Castaneda, O. A., Thongchan, D., Khatri-Chhetri, R., Tsai, S. S., & Wu, H. Y. (2017). Pigeon circovirus infection in disqualified racing pigeons from Taiwan. *Avian Pathol, 46*(4), 359-366. doi:10.1080/03079457.2017.1284305

Kaleta, E. F., & Baldauf, C. (1988). Newcastle disease in free-living and pet birds *Newcastle disease* (pp. 197-246): Springer.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol, 35*(6), 1547-1549. doi:10.1093/molbev/msy096

Ledwoń, A., Bailey, T., O'Donovan, D., Mckeown, S., Lloyd, C., Więckowski, T., . . . Wernery, U. (2011). Prevalence of circovirus and adenovirus in pigeons in Dubai. *Medycyna Weterynaryjna, 67*, 752-756.

Loiko, M. R., Junqueira, D. M., Varela, A. P. M., Tochetto, C., Scheffer, C. M., Lima, D. A., . . . Roehe, P. M. (2018). Columbid circoviruses detected in free ranging pigeons from Southern Brazil: insights on PiCV evolution. *Arch Virol, 163*(11), 3083-3090. doi:10.1007/s00705-018-3990-8

Mankertz, A., Hattermann, K., Ehlers, B., & Soike, D. (2000). Cloning and sequencing of columbid circovirus (coCV), a new circovirus from pigeons. *Arch Virol, 145*(12), 2469-2479. doi:10.1007/s007050070002

Martin, D. P., Varsani, A., Roumagnac, P., Botha, G., Maslamoney, S., Schwab, T., . . . Murrell, B. (2020). RDP5: A computer program for analysing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evolution*.

McCowan, C., Crameri, S., Kocak, A., Shan, S., Fegan, M., Forshaw, D., . . . Wang, J. (2018). A novel group A rotavirus associated with acute illness and hepatic necrosis in pigeons (Columba livia), in Australia. *PLoS One, 13*(9), e0203853. doi:10.1371/journal.pone.0203853

Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS One, 9*(9), e108277. doi:10.1371/journal.pone.0108277

Omar, M. A., Hassan, F. A., & Shahin, S. E. (2017). Economic Assessment of Various Levels of Protein and Energy in Pigeon Squabs Diet. *Zagazig Veterinary Journal, 45*(Supplementary 1), 20-30.

Phan, T. G., Vo, N. P., Boros, A., Pankovics, P., Reuter, G., Li, O. T., . . . Delwart, E. (2013). The viruses of wild pigeon droppings. *PLoS One, 8*(9), e72787. doi:10.1371/journal.pone.0072787

Proskura, W. S., Kustosz, J., Dybus, A., & Lanckriet, R. (2015). Polymorphism in dopamine receptor D4 gene is associated with pigeon racing performance. *Anim Genet, 46*(5), 586-587.

Sarker, S., Das, S., Ghorashi, S. A., Forwood, J. K., & Raidal, S. R. (2019). Pigeon circoviruses from feral pigeons in Australia demonstrate extensive recombination and genetic admixture with other circoviruses. *Avian Pathol, 48*(6), 512-520. doi:10.1080/03079457.2019.1629391

Stenzel, T., Dziewulska, D., Tykalowski, B., & Koncicki, A. (2020). The Clinical Infection with Pigeon Circovirus (PiCV) Leads to Lymphocyte B Apoptosis But Has No Effect on Lymphocyte T Subpopulation. *Pathogens, 9*(8). doi:10.3390/pathogens9080632

Stenzel, T., & Pestka, D. (2014). Occurrence and genetic diversity of pigeon circovirus strains in Poland. *Acta Vet Hung, 62*(2), 274-283. doi:10.1556/AVet.2014.004

Stenzel, T., Piasecki, T., Chrząstek, K., Julian, L., Muhire, B. M., Golden, M., . . . Varsani, A. (2014). Pigeon circoviruses display patterns of recombination, genomic secondary structure and selection similar to those of beak and feather disease viruses. *Journal of General Virology, 95*(6), 1338-1351.

Stenzel, T. A., Pestka, D., Tykalowski, B., Smialek, M., & Koncicki, A. (2012). Epidemiological investigation of selected pigeon viral infections in Poland. *Vet Rec, 171*(22), 562. doi:10.1136/vr.100932

Stover, B. C., & Muller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics, 11*(1), 7. doi:10.1186/1471-2105-11-7

Teske, L., Rubbenstroth, D., Meixner, M., Liere, K., Bartels, H., & Rautenschlein, S. (2017). Identification of a novel aviadenovirus, designated pigeon adenovirus 2 in domestic pigeons (Columba livia). *Virus Res, 227*, 15-22. doi:10.1016/j.virusres.2016.09.024

Todd, D., Fringuelli, E., Scott, A. N., Borghmans, B. J., Duchatel, J. P., Shivaprasad, H. L., . . . Smyth, J. A. (2008). Sequence comparison of pigeon circoviruses. *Res Vet Sci, 84*(2), 311-319. doi:10.1016/j.rvsc.2007.03.007

Todd, D., Weston, J. H., Soike, D., & Smyth, J. A. (2001). Genome sequence determinations and analyses of novel circoviruses from goose and pigeon. *Virology, 286*(2), 354-362. doi:10.1006/viro.2001.0985

Wan, C., Chen, C., Cheng, L., Shi, S., Fu, G., Liu, R., . . . Huang, Y. (2018). Detection of novel adenovirus in sick pigeons. *J Vet Med Sci, 80*(6), 1025-1028. doi:10.1292/jvms.18-0024

Wang, K. C., Zhuang, Q. Y., Qiu, Y., Wang, T., & Chen, J. M. (2017). Genome sequence characterization of pigeon circoviruses in China. *Virus Res, 233*, 1-7. doi:10.1016/j.virusres.2017.03.007

Woods, L. W., Latimer, K. S., Niagro, F. D., Riddell, C., Crowley, A. M., Anderson, M. L., . . . Nordhausen, R. W. (1994). A retrospective study of circovirus infection in pigeons: nine cases (1986-1993). *J Vet Diagn Invest, 6*(2), 156-164. doi:10.1177/104063879400600205

Zhang, Z., Dai, W., & Dai, D. (2017). Molecular characterization of pigeon torque teno virus (PTTV) in Jiangsu province. *Comput Biol Chem, 69*, 10-18. doi:10.1016/j.compbiolchem.2017.04.012

Chapter 4:

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology, 215*(3), 403-410.

Ball, N., Smyth, J., Weston, J., Borghmans, B., Palya, V., Glavits, R., . . . Todd, D. (2004). Diagnosis of goose circovirus infection in Hungarian geese samples using polymerase chain reaction and dot blot hybridization tests. *Avian Pathology, 33*(1), 51-58.

Banda, A., Galloway-Haskins, R. I., Sandhu, T. S., & Schat, K. A. (2007). Genetic analysis of a duck circovirus detected in commercial Pekin ducks in New York. *Avian diseases, 51*(1), 90-95.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Prjibelski, A. D. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology, 19*(5), 455-477.

Chandler, M., De La Cruz, F., Dyda, F., Hickman, A. B., Moncalian, G., & Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews Microbiology, 11*(8), 525-538.

Darriba, D., Taboada, G., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27. *DOI: https://doi. org/10.1093/bioinformatics/btr088, PMID, 21335321*.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods, 9*(8), 772-772.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., . . . Stiassny, M. L. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews, 81*(2), 163-182.

Guindon, S., Delsuc, F., Dufayard, J.-F., & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML *Bioinformatics for DNA sequence analysis* (pp. 113-137): Springer.

Halami, M., Nieper, H., Müller, H., & Johne, R. (2008). Detection of a novel circovirus in mute swans (Cygnus olor) by using nested broad-spectrum PCR. *Virus research, 132*(1-2), 208-212.

Halvorson, D., Karunakaran, D., Senne, D., Kelleher, C., Bailey, C., Abraham, A., . . . Newman, J. (1983). Epizootiology of avian influenza: simultaneous monitoring of sentinel ducks and turkeys in Minnesota. *Avian diseases*, 77-85.

Hattermann, K., Schmitt, C., Soike, D., & Mankertz, A. (2003). Cloning and sequencing of Duck circovirus (DuCV). *Archives of virology, 148*(12), 2471-2480.

Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *Plos biol, 2*(10), e312.

Ilyina, T. V., & Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic acids research, 20*(13), 3279-3285.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution, 30*(4), 772-780.

Krupovic, M., Varsani, A., Kazlauskas, D., Breitbart, M., Delwart, E., Rosario, K., . . . Zerbini, F. M. (2020). Cressdnaviricota: a virus phylum unifying 7 families of Rep-encoding viruses with single-stranded, circular DNA genomes. *Journal of virology*.

Lima, F. E. S., Cibulski, S. P., Bello, A. G. D., Mayer, F. Q., Witt, A. A., Roehe, P. M., & d'Azevedo, P. A. (2015). A novel chiropteran circovirus genome recovered from a Brazilian insectivorous bat species. *Genome announcements, 3*(6).

Markwell, D. D., & Shortridge, K. F. (1982). Possible waterborne transmission and maintenance of influenza viruses in domestic ducks. *Applied and environmental microbiology, 43*(1), 110-115.

Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PloS one, 9*(9), e108277.

Pass, D., & Perry, R. (1984). The pathology of psittacine beak and feather disease. *Australian Veterinary Journal, 61*(3), 69-74.

Rosario, K., Breitbart, M., Harrach, B., Segalés, J., Delwart, E., Biagini, P., & Varsani, A. (2017). Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Archives of virology, 162*(5), 1447-1463.

Runstadler, J., Happ, G., Slemons, R., Sheng, Z.-M., Gundlach, N., Petrula, M., . . . Modrell, A. (2007). Using RRT-PCR analysis and virus isolation to determine the prevalence of avian influenza virus infections in ducks at Minto Flats State Game Refuge, Alaska, during August 2005. *Archives of virology, 152*(10), 1901-1910.

Segalés, J. (2012). Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis. *Virus research, 164*(1-2), 10-19.

Siembieda, J. L., Johnson, C. K., Cardona, C., Anchell, N., Dao, N., Reisen, W., & Boyce, W. (2010). Influenza A viruses in wild birds of the Pacific flyway, 2005–2008. *Vector-Borne and Zoonotic Diseases, 10*(8), 793-800.

Stöver, B. C., & Müller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC bioinformatics, 11*(1), 7.

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society, 29*(1), 344-358.

Todd, D., Weston, J., Soike, D., & Smyth, J. (2001). Genome sequence determinations and analyses of novel circoviruses from goose and pigeon. *Virology, 286*(2), 354-362.

Wallensten, A., Munster, V. J., Latorre-Margalef, N., Brytting, M., Elmberg, J., Fouchier, R. A., . . . Lundkvist, Å. (2007). Surveillance of influenza virus A in migratory waterfowl in Northern Europe. *Emerging infectious diseases, 13*(3), 404.

Woods, L. W., Latimer, K. S., Niagro, F. D., Riddell, C., Crowley, A. M., Anderson, M. L., . . . Nordhausen, R. W. (1994). A retrospective study of circovirus infection in pigeons: nine cases (1986–1993). *Journal of Veterinary Diagnostic Investigation, 6*(2), 156-164.

Wu, Z., Ren, X., Yang, L., Hu, Y., Yang, J., He, G., . . . Du, J. (2012). Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *Journal of virology, 86*(20), 10999-11012.

APPENDIX A
AUTHOR'S PERMISSION

The research presented in chapter 2 has been published in the journal *Microbiology Resource Announcements*. Anthony Khalifeh, Simona Kraberger, Daria Dziewulska, Tomasz Stenzel, Arvind Varsani. All co-authors have granted permission for this work to be included in this dissertation.

The research presented in chapter 3 was submitted to the journal *Viruses* and is under review. Anthony Khalifeh, Simona Kraberger, Daria Dziewulska, Tomasz Stenzel, Arvind Varsani. All co-authors have granted permission for this work to be included in this dissertation.

All other associated authors not yet listed provided permission for the use of the data in this dissertation.