Identification of CRESS DNA viruses in faeces of Pacific flying foxes in the Tongan archipelago

A thesis submitted in partial fulfilment of the requirements for the Degree of
Master of Science
at the
University of Canterbury, New Zealand

Maketalena F. Male

2016

Table of contents

Acknowledgements

This thesis is possible because of the support from my scholarships, an Arthington Fund Scholarship from Trinity College (Cambridge, UK) and a Government of Tonga Scholarship. Thanks to Paula Jameson for helping me secure this funding.

I would also like to show my appreciation and gratitude by acknowledging the following amazing people for all their help throughout my thesis research.

To Arvind Varsani, thank you for being the best supervisor I could ever ask for. His unending support, guidance and patience made this thesis happen.

To my co-supervisor, Simona Kraberger, who guided me in lab and computational work, thank you for having the time to teach me and introducing me to the field of research.

To my associate-supervisor, David Collings, who has helped throughout my time here at the University of Canterbury, thank you for always being cheerful and for continuously making biology interesting.

To Matt Walters and Daisy Stainton, thank you so much for all your help. Daisy has been very encouraging for as long as I can remember.

To Viliami Kami and the Ministry of Agriculture, Food, Forests & Fisheries in Tonga, I really appreciate the help during my sample collection.

Cheers to all my office and lab mates especially Simona, Daisy, Dorien, Marta and Anisha. My time at the School of Biological Sciences has been full of joy and awesome memories because of you all.

Words cannot express my deep appreciation and gratefulness towards my Tongan and US family. To my parents Vaea and Loukinikini, my siblings, my best-friend Tongaliuaki and the Edwards, thank you for always being there for me.

This thesis is dedicated to my father Vaea Tangitau ‘Akau’ola who has gone but never forgotten. Thanks for always believing in me.

Abstract

Viruses with circular single-stranded DNA (ssDNA) genomes are the smallest pathogens known to infect various organisms. Due to advances in high-throughput sequencing technologies, the diversity of circular replication associated protein encoding single-stranded (CRESS) DNA viruses is beginning to unravel. Viral metagenomic studies have demonstrated that animal faecal matter harbours a high viral diversity and therefore can potentially be used to explore viruses within ecosystems. Faecal matter may contain viruses shed by the infected animal or those that are associated with its diet and the environment. Besides capturing the viral diversity, faecal sampling is a non-invasive to the animal hence can be used easily for viral surveillance in ecosystems.

A limited amount of work has been done on CRESS DNA viruses circulating in the Pacific Islands of Tonga. Prior to this study, only six species of CRESS DNA viruses had been identified. As part of a continuing effort to determine the diversity of CRESS DNA viruses, I sampled *Pteropus tonganus* faeces. *P. tonganus*, also known as the Pacific flying fox, is the most widespread bat species in the Pacific and is the only bat species found in the Tongan archipelago. Pacific flying foxes roost in trees and are frugivores.

This thesis research was carried out to identify CRESS DNA viruses that are associated with Pacific flying fox faeces in Tonga. Faecal samples were collected from four *P. tonganus* roosting sites (Ha'ateiho ('Atele), Lapaha (Takuilau), Ha'avakatolo and Kolovai) located in Tongatapu the main island of Tonga in 2014 and 2015. A next-generation sequencing informed approach was used to recover complete CRESS DNA viral genomes. In total, five novel cycloviruses (three species), 25 novel gemycircularviruses (13 species), 17 unclassified novel CRESS DNA viruses (15 species), a putative multicomponent virus (three cognate molecules) and two circular DNA molecules, were recovered. A number of viruses were identified in more than one sampling site in Tonga, suggesting these viruses have a broad distribution across the island amongst the Pacific flying fox colonies. Several species were identified in both 2014 and 2015 suggesting these viruses are persistently associated with faecal matter of Pacific flying foxes. The data obtained from this study has significantly expanded the knowledge of CRESS DNA viruses that are circulating in Tonga.

Deputy Vice-Chancellor's Office
Postgraduate Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 2

Male, M. F., Kraberger, S., Stainton, D., Kami, V. & Varsani, A. (2016). Cycloviruses, gemycircularviruses and other novel replication-associated protein encoding circular viruses in Pacific flying fox (Pteropus tonganus) faeces. Infection, genetics and evolution. In Press

Please detail the nature and extent (%) of contribution by the candidate:

The field work was undertaken by Maketalena Male with the help of Viliami Kami.

All the molecular work was undertaken by Maketalena Male. She worked with Arvind Varsani and Simona Kraberger to analyse the data. Arvind Varsani double checked all the bioinformatics analysis. Daisy Stainton helped with checking properties of the multicomponent virus identified in this study.

Maketalena Male wrote the manuscript and the rest of the authors provided comments / feedback to improve it.

Maketalena Male contribution: 85%

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifys that:

- The above statement correctly reflects the nature and extent of the PhD candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: Arvind Varsani Signature:  Date: *20 Jan 2016*

Chapter 1

Introduction

Contents

1.1 Classification of viruses

Viruses are documented to be one of the most abundant biological entities on Earth with members which are able to infect all the three domains of life and are found in all types of environments (Breitbart & Rohwer, 2005; Edwards & Rohwer, 2005; King *et al.*, 2012). The continued expansion of knowledge on viruses has shown that they have played an important role in the evolution of life, from introducing new functions into the genomes of various organisms to facilitating gene transfer and controlling microbial populations (Forterre & Prangishvili, 2009; Rohwer *et al.*, 2009; Suttle, 2007). With the rapid discovery of novel viruses, largely attributed to viral metagenomics studies, it has become evident that very little is known about the breadth of viral diversity present on Earth (Edwards & Rohwer, 2005).

In order to place the viruses which are explored in this thesis into the larger context of all viruses, it is important to understand how viruses are classified. The Baltimore classification (Baltimore, 1971), broadly classifies viruses into seven groups based on their genome type i.e RNA or DNA genomes that are either single or double stranded, the polarity of the single stranded nucleic acid genomes and their mode of replication (King *et al.*, 2012). The seven groups of viruses include double-stranded DNA (dsDNA) viruses, single-stranded DNA (ssDNA) viruses, dsRNA viruses, positive-sense (+)ssRNA viruses, negative-sense (-)ssRNA viruses, retro-transcribing ssRNA-RT viruses and retro-transcribing dsDNA-RT viruses (King *et al.*, 2012) (Table 1.1).

The international committee for virus taxonomy (ICTV) was then established to assess the virus taxonomy in the context of the Baltimore system, based on current knowledge of virus diversity and classification (King *et al.*, 2012). The taxonomic classification of viruses expands to include several characteristics such as virion structure and whether the virions are enveloped or not, the genome types, genome organisation, host range, tissue tropism and the mechanism of viral replication (King *et al.*, 2012). Following this, each group were simplified into order, family, genus and species.

Table 1.1: Overview of the Baltimore classification of viruses, including genome type, mRNA synthesis and an example from each group.

Group	Genome type	mRNA synthesis	Examples
I	dsDNA	+/-dsDNA → +mRNA	Adenoviruses
II	ssDNA	+ssDNA → +/-dsDNA → +mRNA	Circoviruses
III	dsRNA	+/-dsRNA → +mRNA	Reoviruses
IV	(+)ssRNA	+ssRNA → -ssRNA → +mRNA	Togaviruses
V	(-)ssRNA	-ssRNA → +mRNA	Rhabdoviruses
VI	ssRNA-RT	+ssRNA → -ssDNA → +/-dsDNA → +mRNA	Retroviruses
VII	dsDNA-RT	+/-dsDNA → +ssRNA → -ssDNA → +/-dsDNA → +mRNA	Hepadnaviruses

1.2 Diversity and classification of single-stranded DNA viruses

Knowledge of ssDNA viral diversity has increased significantly in the last decade. This has been attributed partly to advances in molecular techniques and sequencing technologies. Viral metagenomics has also enabled the discovery of novel ssDNA viruses in a wide range of plants, animals, fungi, bacteria, and environmental samples.

SsDNA viruses are classified into nine families i.e. *Inoviridae* and *Microviridae* whose members infect prokaryotes, *Anelloviridae*, *Bidnaviridae*, *Circoviridae*, *Geminiviridae*, *Nanoviridae* and *Parvoviridae* whose members infect eukaryotes and *Spiraviridae* whose members infect archaea (Table 1.2) (Adams & Carstens, 2012; King *et al.*, 2012; Mochizuki *et al.*, 2012). Most of the families of ssDNA viruses are encapsidated in icosahedral virions except for inoviruses which consist of viruses with a filamentous or rod-shaped morphology, spiraviruses with coil shaped particle and geminiviruses with twinned icosahedral particles (King *et al.*, 2012) (Table 1.2). The majority of the members of ssDNA viral families have circular DNA genomes with the exception of *Parvoviridae* and *Bidnaviridae* families which have linear ssDNA genomes (King *et al.*, 2012) (Table 1.2). The continued discovery of novel ssDNA viruses indicates that their diversity has been grossly underestimated and it is expected that classification will continue to change with the discovery of additional novel ssDNA viruses. An example of such change has been shown by the proposal to create the new family Genomoviridae for the proposed genus gemycircularvirus (see Genomoviridae proposal at ICTV).

Table 1.2: Overview of the accepted nine families of single-stranded DNA viruses according to the ICTV.

Host organisms	ssDNA virus families	Capsid morphology	DNA genome	Genome size (kb)	Host organism genome structure
Bacteria	*Inoviridae*	Filamentous or rod-shaped	Circular	4.5 - 12.4	Monopartite
	Microviridae	Isosahedral	Circular	4.4 - 5.3	Monopartite
Animals	*Anelloviridae*	Isosahedral	Circular	2.8 - 3.9	Monopartite
	Circoviridae *	Isosahedral	Circular	1.7 - 2.3	Monopartite
Animals and insects	*Parvoviridae*	Isosahedral	Linear	4 - 6.3	Monopartite
Insects	*Bidnaviridae*	Isosahedral	Linear	13	Multipartite
Plants	*Geminiviridae*	Twinned isosahedral	Circular	2.7 - 5.4	Monopartite and bipartite
	Nanoviridae	Icosahedral	Circular	6.4 - 8	Multipartite
Archaea	*Spiraviridae*	Coil shaped	Circular	24.8	Monopartite

* Identified in this study

1.2.1 Prokaryote infecting ssDNA viruses

1.2.1.1 *Inoviridae*

Members of the *Inoviridae* family contain a circular, positive sense ssDNA genome with a genome size of 4.5 - 12.4 kilobases (kb) and all members are known to infect prokaryotes (King *et al.*, 2012). This family is divided into two genera, *Inovirus* and *Plectrovirus*. Members of the *Inovirus* genus (e.g. *Enterobacteria phage*, *Pseudomonas phage*, *Vibrio phage* and *Xanthomonas phage*) infects gram-negative and gram-positive bacteria (King *et al.*, 2012). Members of the *Plectrovirus* genus (e.g. *Acholeplasma phage* and *Spiroplasma phage*) are known to infect mycoplasma (King *et al.*, 2012). Many environmental sampling methods have detected inoviruses. For example, inoviruses have been recovered from alkaline hot spring, soil, fermented foods, raw sewage, lagoon wastewater and Arctic sea ice (Alhamlan *et al.*, 2013; Cantalupo *et al.*, 2011; Park *et al.*, 2011; Swanson *et al.*, 2009; Yu *et al.*, 2006; Yu *et al.*, 2014). Furthermore, inoviruses have also been detected in pharyngeal and rectal swabs of insectivorous bats (Wu *et al.*, 2012; Wu *et al.*, 2015).

1.2.1.2 *Microviridae*

The *Microviridae* family are non-enveloped with small icosahedral morphology with circular ssDNA genomes (King *et al.*, 2012). Members are divided into one genus, *Microvirus*, and one sub-family, *Gokushovirinae* (King *et al.*, 2012). The genus *Microvirus* exclusively infects *Enterobacteria* spp (King *et al.*, 2012). Members of the sub-family *Gokushovirinae* are known to infect obligate intracellular parasites such as *Chlamydia* spp., *Bdellovibrio* spp. and *Spiroplasma* spp. (Brentlinger *et al.*, 2002). The virion morphology, genome

organization, genome identities and host lifestyle of the members of the sub-family *Gokushovirinae* are different from the members of *Microvirus* (King *et al.*, 2012). Microvirus-like sequences have been detected in a variety of samples including marine environments, soil, lakes, peatlands, dragonflies, human, turkey, methane seep sediments and bats (Bryson *et al.*, 2015; Labonte & Suttle, 2013; Li *et al.*, 2010b; Minot *et al.*, 2011; Quaiser *et al.*, 2015; Reavy *et al.*, 2015; Rosario *et al.*, 2012a; Roux *et al.*, 2012; Tucker *et al.*, 2011; Zhong *et al.*, 2015; Zsak *et al.*, 2011). Specifically, the microvirus phages identified in bats were recovered from their faeces and were associated with the food they eat (Li *et al.*, 2010b).

1.2.2 Eukaryote infecting ssDNA viruses

1.2.2.1 *Anelloviridae*

Anelloviridae family consist of small non-enveloped viruses that infect a wide range of mammalian species (Biagini *et al.*, 2006). Their genomes are negative-sense circular ssDNA that have an untranslated region (UTR) and at least two major open reading frames (ORFs) (Biagini *et al.*, 2006; Rosario *et al.*, 2012b). Currently there are nine genera assigned to the *Anelloviridae* family (King *et al.*, 2012). The genus *Gyrovirus* was originally from another ssDNA family but has recently been assigned as a genus within the *Anelloviridae* family (see *Gyrovirus* proposal at ICTV). Members of the *Anelloviridae* family have been recovered from various mammalian and reptiles tissues (Biagini *et al.*, 2007; Bouzari & Salmanizadeh, 2015; Burian *et al.*, 2011; Cibulski *et al.*, 2014; Huang *et al.*, 2010; Ng *et al.*, 2009; Ng *et al.*, 2011b; Ninomiya *et al.*, 2009; Nishiyama *et al.*, 2014; Okamoto, 2009; Young *et al.*, 2015). Currently, there are no reports of annelloviruses in bat faeces, the only Annellovirus detected in bats was from organs of Brazilian free-tailed bats (Cibulski *et al.*, 2014).

1.2.2.2 *Bidnaviridae*

The newly assigned family *Bidnaviridae* has only one genus, *Bidensovirus*, which houses a species *Bombyx mori bidensovirus* (BmBDV) that was recently moved from the family *Parvoviridae* (Adams & Carstens, 2012). BmBDV is known to cause diseases in the silkworm *Bombyx mori* by infecting the columnar cells of the midgut epithelium where the results can be fatal (Hu *et al.*, 2013). BmBDV is a non-enveloped spherical virus with linear segmented ssDNA of 6.5 kb (DNA1, VD1) and 6 kb (DNA2, VD2) (Hu *et al.*, 2013). The two DNA molecules are packaged in two different capsids and the replication mechanism of

BmBDV is yet to be uncovered (Hu *et al.*, 2013). BmBDV encodes a putative protein-primed DNA polymerase (Hu *et al.*, 2013). Note that this family has only been found to infect insect cells but has not been identified in any environmental samples such as faecal matter.

1.2.2.3 *Circoviridae*

Members of the family *Circoviridae* are found to infect various animals. The members of this family have an ambisense genome organization consisting of ~1.7 - 2.3 kb and their virions have icosahedral structure (King *et al.*, 2012). The family *Circoviridae* currently have one genus i.e. *Circovirus*, however, there is also an additional genus that has been recently proposed which is the cyclovirus (see cyclovirus proposal at ICTV). The majority of circoviruses have been found to be infecting a wide range of avian species, fish and a few mammals (Allan & Ellis, 2000; He *et al.*, 2013; Kapoor *et al.*, 2012; Lorincz *et al.*, 2011; Todd, 2004; Wu *et al.*, 2012; Wu *et al.*, 2015). Cycloviruses on the other hand have been recovered from various mammals and insect samples (Dayaram *et al.*, 2013b; Ge *et al.*, 2011; Li *et al.*, 2011; Li *et al.*, 2010b; Lima *et al.*, 2015; Padilla-Rodriguez *et al.*, 2013; Phan *et al.*, 2014; Rosario *et al.*, 2012a; Rosario *et al.*, 2011; Smits *et al.*, 2013; van Doorn *et al.*, 2013; Victoria *et al.*, 2009; Wu *et al.*, 2015). In particular, circoviruses and cycloviruses have been detected in bat samples including pharyngeal and rectal swabs, muscle, stomach contents and faeces (Ge *et al.*, 2011; He *et al.*, 2013; Li *et al.*, 2010a; Li *et al.*, 2011; Lima *et al.*, 2015; Wu *et al.*, 2012; Wu *et al.*, 2015).

1.2.2.4 *Geminiviridae*

Geminiviruses are plant infecting viruses that have circular ssDNA genomes ranging in size from ~2.7 - 5.4 kb and the genomes are encapsidated in twinned icosahedral shaped particles or geminate virions (King *et al.*, 2012). Members in this family are currently classified into seven genera, as reviewed in (King *et al.*, 2012; Varsani *et al.*, 2014a; Varsani *et al.*, 2014b). The majority of geminiviruses have monopartite genomes which consist of one circular ssDNA component. Members of the *Begomovirus* genus on the other hand can either be monopartite (a single component) or bipartite (two components known as DNA-A and DNA-B) genomes (Briddon *et al.*, 2010; Brown *et al.*, 2015). Furthermore, some begomoviruses are often associated with satellite DNA molecules known as alphasatellites and betasatellites, with betasatellites being significant in inducing pathogenicity of begomoviruses (Nawaz-Ul-Rehman *et al.*, 2010). Geminiviruses are transmitted by whitefly, leafhoppers and tree

hoppers (King *et al.*, 2012). A large number of geminiviruses have been identified in numerous plants (Brown *et al.*, 2015; Fauquet *et al.*, 2003; King *et al.*, 2012; Muhire *et al.*, 2014; Varsani *et al.*, 2014a; Varsani *et al.*, 2014b). To date, members of this family have never been identified in bat faeces.

1.2.2.5 *Nanoviridae*

The family *Nanoviridae* consist of plant infecting viruses that have multi-partite or multi-component circular ssDNA genomes (6-8 segments), each ~1 kb (King *et al.*, 2012). *Nanovirus* and *Babuvirus* are the two genera in this family, both are vectored by aphids and are known to infect dicotyledonous and monocotyledonous plants, respectively (Mandal, 2010). In nanoviruses and babuviruses, the circular ssDNA molecules each encode a single ORF and each DNA molecule is packaged into separate icosahedral particles. Within the genomes of all the components of nanoviruses and babuviruses, there are two common regions that are common across the integral components and they are the common region stem-loop (CR-SL) and the common region major (CR-M) (Figure 1.1). To date, eight and three assigned species of nanoviruses and babuviruses, respectively, have been identified in various plants (Abraham *et al.*, 2012; Burns *et al.*, 1995; Chu & Helms, 1988; Grigoras *et al.*, 2014; Grigoras *et al.*, 2010a; Grigoras *et al.*, 2009; Katul *et al.*, 1998; Kumari *et al.*, 2010; Mandal *et al.*, 2013; Sano *et al.*, 1998; Sharman *et al.*, 2008). Like the geminiviruses, members of the nanoviruses have never been identified in any environmental sampling. As some bat species are frugivores, there is potential for plant infecting viruses to be present in their faeces.

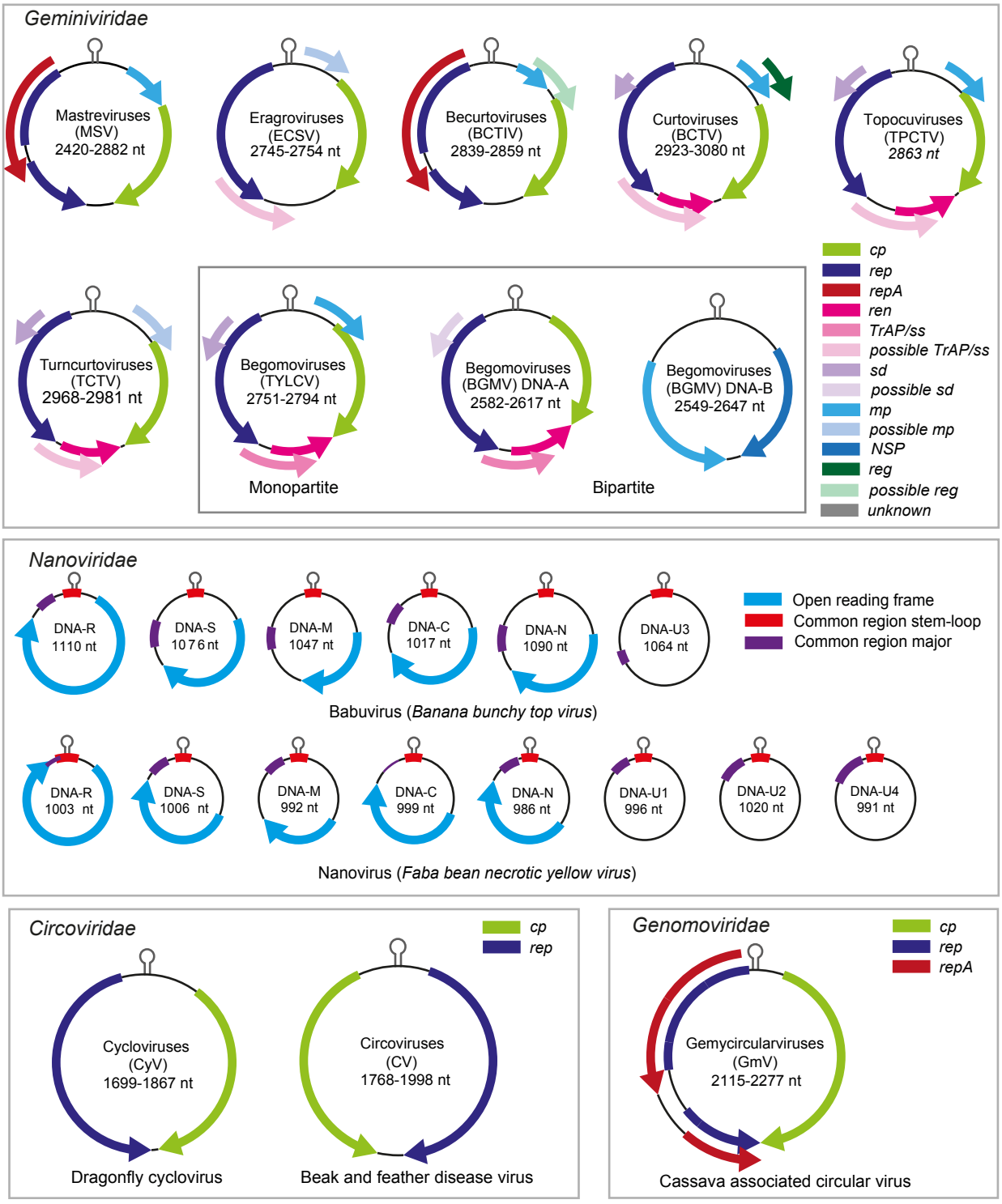


Figure 1.1: An illustration of the genome organization of representative viruses from the *Geminiviridae*, *Nanoviridae*, *Circoviridae* and proposed Genomoviridae families. Members of the *Nanoviridae* family have common regions and the components are known as DNA-R (encoding replication associated protein), DNA-S (encoding capsid protein), DNA-M (encoding movement protein), DNA-C (encoding cycle link protein), DNA-N (encoding nuclear shuttle protein) and DNA-U1, U2, U3 and U4 (encoding ORFs of unknown function).

1.2.2.6 *Parvoviridae*

Members of the family *Parvoviridae* have small isometric particles that encapsidate linear ssDNA genomes of approximately 4 - 6 kb (Tattersall, 2006). Parvoviruses infect a wide range of hosts and hence are divided into two sub-families: the *Parvovirinae* which infect vertebrates and *Densovirinae* containing viruses which infect arthropods (Cotmore *et al.*, 2014). The sub-family *Parvovirinae* contain eight genera and the sub-family *Densovirinae* has five genera (Cotmore *et al.*, 2014). Few members of parvoviruses are accountable for many pathogenic diseases in animals and humans (Allander & Andersson, 2012; Delwart & Jones, 2014; Steinel *et al.*, 2001). Interestingly, studies on bat faecal and anal swab samples reported viral sequences related to those in two sub-families of *Parvoviridae* (Ge *et al.*, 2012; Wu *et al.*, 2015).

1.2.3 Archaea infecting ssDNA viruses

1.2.3.1 *Spiraviridae*

The family *Spiraviridae* has recently been established and has one genus, *Spiravirus*. The only assigned species to the *Spiravirus* is the *Aeropyrum coil-shaped virus* (ACV). The genomic properties and morphology of ACV together with its exceptionally large genome size of 24.8 kb makes it a unique virus amongst all ssDNA viruses (Mochizuki *et al.*, 2012). ACV has been isolated from an environmental sample of the hyperthermophilic archaea *Aeropyrum pernix* that was collected from Yamagawa hot spring of 104 °C in Japan (Mochizuki *et al.*, 2010). It is noteworthy that ACV is the only archaea infecting virus with a circular ssDNA genome (Mochizuki *et al.*, 2012). It has a linear, hollow coil-shaped and non-enveloped particle structure formed from a coiling fibre, which consists of two intertwining halves of a single circular nucleoprotein (Mochizuki *et al.*, 2012). There are no reports of ACV from any other environmental sample.

1.2.4 The proposed family Genomoviridae

The proposed family of Genomoviridae has one proposed genus named gemycircularvirus. Members of the proposed genus gemycircularvirus have ~2 kb circular ssDNA genomes which are ambisense encoding a capsid protein (CP) in the virion sense and a replication-associated protein (Rep) in the complementary sense. The Reps of gemycircularviruses share similarities with Reps of some geminiviruses and mycovirus-like sequences (Kraberger *et al.*,

2015). Of all the previously identified putative gemycirularviruses, only Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1) has been associated with a known host, *Sclerotinia sclerotiorum*, inducing hypovirulence (Yu *et al.*, 2010). Interestingly, SsHADV-1 is the only ssDNA known to infect fungi (Jiang *et al.*, 2013). Numerous metagenomics studies of different environmental samples, various animal tissues and plants have identified sequences that share similarities with SsHADV-1 (see Genomoviridae proposal at ICTV). To date, there has been no gemycircularvirus identified in bat faeces, however, three species have been found in pharyngeal and rectal swabs of insectivores bats (Wu *et al.*, 2015).

1.3 Circular replication associated protein encoding ssDNA viruses

All ssDNA viruses with circular genomes that encode a well conserved replication-associated protein (Rep) are also refer to as circular Rep-encoding ssDNA (CRESS DNA) viruses (Rosario *et al.*, 2012a). The three well established families that are part of the CRESS DNA viruses are *Circoviridae*, *Geminiviridae* and *Nanoviridae* (Figure 1.1). The proposed family Genomoviridae is also part of CRESS DNA viruses (Figure 1.1). In addition to these are varieties of CRESS DNA viruses which have been identified from various sources that have not been assigned to specific taxons. Various genome organisations have been noted for these (Rosario *et al.*, 2012b) and these are summarised in Figure 1.2.

CRESS DNA viruses replicate through a mechanism known as rolling circle replication (RCR) (Gutierrez, 1999; Hanley-Bowdoin *et al.*, 1999; Stenlund, 2003; Timchenko *et al.*, 1999) as summarised in (Figure 1.3) and as reviewed in Gutierrez (1999) and Rosario *et al.* (2012b). Firstly, the ssDNA genome undergoes a conversion from single-stranded to a double-stranded replicative form using the host factors. In the initiation stage, the Rep is responsible for initiating replication by recognising and binding to the replicative form close to the origin of replication (*ori*) leading to a partial melt in the dsDNA strand (Gutierrez, 1999). The Rep then nicks a specific recognition of the conserved nonanucleotide sequence (Hafner *et al.*, 1997). The nick creates a primer for leading-strand synthesis on the 3'-OH overhangs which then are identified by the host polymerases (Gutierrez, 1999). The 5' end is still covalently attached to the Rep and the elongation stage then proceeds. Through replication of the circular molecule, the template ssDNA strand is replaced by the newly synthesised ssDNA strand. Lastly, during the termination stage the Rep facilitates the joining of the newly synthesized ssDNA strand while the template strand is released and the RCR

cycle is completed (Hafner *et al.*, 1997; Laufs *et al.*, 1995). It is important to note that the overview of RCR is from experimental data collected from studies on the replication of geminiviruses (Gutierrez, 1999; Hanley-Bowdoin *et al.*, 1999) and nanoviruses (Hafner *et al.*, 1997; Timchenko *et al.*, 1999).

Figure 1.2: An illustration showing the eight different genome organisation of CRESS DNA molecules. Each illustration shows single or multiple ORFs, ambisense or unisense genome arrangement and the positioning of the stem-loop. The arrows represent the direction of the ORFs showing whether they read towards the stem-loop, away from the stem-loop, clockwise or anti-clockwise. Modified from Rosario *et al.* (2012b).

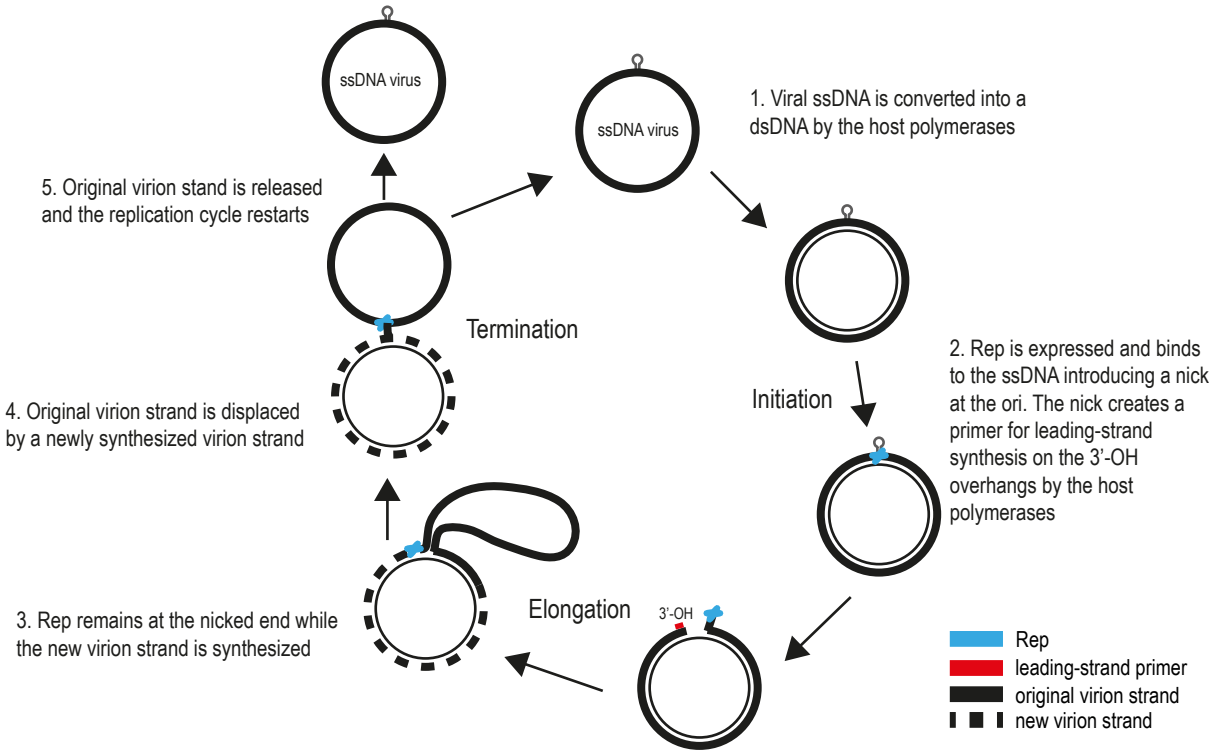


Figure 1.3: An cartoon illustration summarising rolling circle replication (RCR) of CRESS DNA viruses, as modified from Rosario *et al.* (2012b). Upon entering the host cell nucleus, ssDNA is converted to dsDNA replicative form by the host polymerases (step 1). The RCR process is then preceded beginning with initiation (step 2) followed by elongation (step 3 and 4) and ending with termination (Step 5).

1.4 Evolution of CRESS DNA viruses

1.4.1 Mutation rates

In general, it has been assumed that RNA viruses evolve at a fast rate compared to DNA viruses due to the fact that they are relying on the low fidelity RNA polymerases for replication (Domingo, 1994; Duffy *et al.*, 2008; Hanley-Bowdoin *et al.*, 1999; Rojas *et al.*, 2005). The estimation of nucleotide substitution rate has been used to show that RNA viruses evolve faster than DNA viruses. However, numerous studies have shown that CRESS DNA viruses have relatively high mutation rates and are evolving at rates similar to RNA viruses (De Bruyn *et al.*, 2012; Duffy & Holmes, 2008; Duffy *et al.*, 2008; Firth *et al.*, 2009; Ge *et al.*, 2007; Grigoras *et al.*, 2010b; Harkins *et al.*, 2009; Kraberger *et al.*, 2013; Shackelton *et al.*, 2005). The overall substitution rates for nearly all RNA viruses fall in the range of 10^{-2} to 10^{-5} nucleotide substitution per site, per year (subs/site/year), with most RNA viruses having rates within one order of magnitude of 1×10^{-3} subs/site/year (Duffy *et al.*, 2008). Studies of ssDNA viruses have shown high substitution rates, for example *Porcine circoviruses* type 2 have an estimated nucleotide substitution rate of 1.2×10^{-3} subs/site/year (Firth *et al.*, 2009), nanovirus *Faba bean necrotic yellows virus* (FBNYV) have 1.78×10^{-3} subs/site/year (Grigoras *et al.*, 2010b) and begomovirus *Tomato yellow leaf curl virus* with an estimated mean rate of 2.88×10^{-4} subs/site/year for the full genome (Duffy & Holmes, 2008).

1.4.2 Recombination

Recombination facilitates rapid exploitation of sequence space and thus diversity. Recombination is the swapping of genetic material from one virus genome/component to another when both co-infect the same cell. See Martin *et al.* (2011a) for a comprehensive review of recombination in eukaryotic ssDNA viruses. Evidence of genetic recombination have been identified in various CRESS DNA viruses including circoviruses (Csagola *et al.*, 2006; Heath *et al.*, 2004; Julian *et al.*, 2013; Stenzel *et al.*, 2014), geminiviruses (Amin *et al.*, 2006; Kraberger *et al.*, 2013; Martin *et al.*, 2011b; Owor *et al.*, 2007; Padidam *et al.*, 1999; Saunders *et al.*, 2002; Varsani *et al.*, 2009a; Varsani *et al.*, 2009b) and nanoviruses (Grigoras *et al.*, 2014; Hu *et al.*, 2007; Hughes, 2004; Islam *et al.*, 2010; Savory & Ramakrishnan, 2014; Stainton *et al.*, 2015a; Stainton *et al.*, 2012). Evidence of recombination has been found in some CRESS DNA viruses, except nanoviruses, to mostly occur at the interface between the CP and the short intergenic region (SIR) and also at the long intergenic region (LIR) near the *v-ori* with lower rates of recombination within the encoding ORFs (Lefeuvre

et al., 2009). Targeted studies are required to determine the exact mechanism of recombination but one mechanism that is thought to occur is the displacement of a replicating strand. This is due to the disruption between the replication enzyme complexes and transcription causing premature displacement followed by reattachment to a different template strand resulting in a recombinant genome (Martin *et al.*, 2011a).

1.4.3 Reassortment

Multi-component viruses such as nanoviruses and geminiviruses are also able to increase their genetic diversity through reassortment also known as pseudo-recombination. Reassortment is the swapping of whole components within species. The separately packaged components are required to be present in one cell in order for reassortment to occur. Evidence of reassortment has been identified in a number of studies of geminiviruses (Chen *et al.*, 2009; Idris & Brown, 2004; Pita *et al.*, 2001) and nanoviruses (Grigoras *et al.*, 2014; Hu *et al.*, 2007; Savory & Ramakrishnan, 2014; Stainton *et al.*, 2015a; Stainton *et al.*, 2012; Yu *et al.*, 2012). In order to detect reassortment events it is preferable that either full genomes or at least more than one component from an isolate is available.

1.5 Approaches to identifying novel CRESS DNA viruses

Viral metagenomics is an unbiased approach for identifying viral nucleic acid sequences in various biological samples without '*a priori*' knowledge of the viral types present (Delwart, 2007; Edwards & Rohwer, 2005). Combined with rolling circle amplification (RCA) using Phi29 DNA polymerase for enrichment of circular DNA molecules, metagenomic approaches have been used in identifying circular CRESS DNA viruses from a range of sample types. Circular DNA genomes have been identified using metagenomics analysis which share similarities to the Reps encoded by members of *Circoviridae*, *Geminiviridae* and *Nanoviridae*, however, these do not fall within these taxons and therefore represent novel and diverse CRESS DNA viruses.

There are conserved motifs in the Reps of all CRESS DNA viruses (Figure 1.4) and these are essential for viral replication through RCR (Rosario *et al.*, 2012b). These conserved motifs are divided into two main categories based on roles, the RCR motifs I, II and III and superfamily 3 (SF3) helicase motifs known as Walker-A, Walker-B and Motif C, as reviewed in Rosario *et al.* (2012b) (Figure 1.4). The three RCR motifs are thought to be involved in recognition and nicking of the DNA for replication as reviewed in Rosario *et al.* (2012b).

Additionally, a conserved motif known as geminivirus Rep sequence (GRS) domain which is located downstream of RCR motif II has been identified in geminiviruses and gemycircularviruses (Kraberger *et al.*, 2015; Nash *et al.*, 2011; Rosario *et al.*, 2012b).

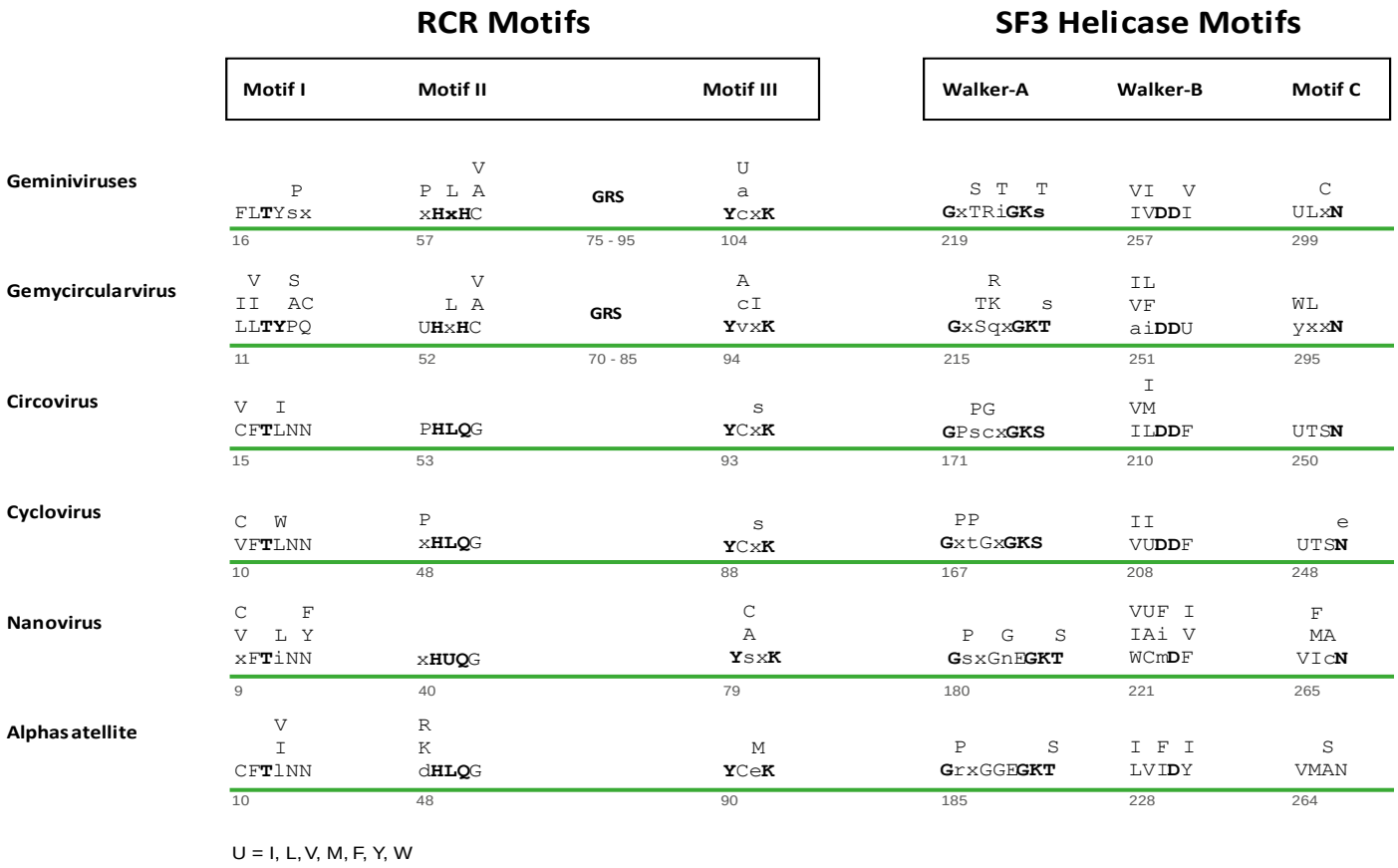


Figure 1.4: A summary of the conserved RCR and SF3 helicase motifs in the replication associated proteins (Reps) of CRESS DNA viruses and alphasatellite, as modified from Rosario *et al.* (2012b). The conserved residues are shaded, uppercase letters are residues that frequently appeared in the analysed sequences and lowercase letters are less frequent residues. If more than four residues were observed in a given position then the letter 'U' or 'x' will be used instead of individual amino acids. 'U' represents bulky hydrophobic amino acids and 'x' for any type. The amino acid numbers that positioned below each motif are derived from representative species of each group, starting from geminivirus to alphasatellite are represented by tomato golden mosaic virus (NC_001507), sewage associated gemycircularvirus-3 (KJ547643), porcine circovirus 1 (NC_001792), PK5222 cyclovirus (GQ404846), *Faba bean necrotic yellows virus* (NC_003560) and *Ageratum conyzoides* alphasatellite. Adapted from (Rosario *et al.*, 2012b).

1.6 Next-generation sequencing platforms

Next-generation sequencing (NGS) or high-throughput sequencing has been used in various studies (Metzker, 2010). Specifically, NGS represent a sequence independent sequencing technique that allows for sequencing of DNA templates without '*a priori*' knowledge of a DNA sequence (Metzker, 2010). The introduction of NGS technology has made large-scale metagenomics analysis more cost effective and importantly, a large number of sequence reads can be generated in a short amount of time (Adams *et al.*, 2009; Shendure & Ji, 2008).

The NGS platforms that are most frequently used in metagenomics studies are Roche 454GS FLX and GS Junior, Illumina HiSeq and MiSeq, SOLiD and Ion Torrent. In addition to this single molecule sequencing platforms such as nanopore and PacBio are bound to play a significant role in future metagenomics. Each platform is different in terms of the chemistry, however, all but PacBio and Nanopore require the preparation of a fragmented DNA library (Shendure & Ji, 2008). Each platform has its own strengths and weaknesses and with constant improvements over the past decade, there is no doubt that sequencing technology will continue to advance for the better.

1.6.1 Roche 454 platform

The Roche 454 system was one of the first NGS platforms. The system relies on the fact that a library of small DNA fragments are ligated to specific adapters that allows for the binding of DNA fragment to specific beads, as reviewed in Ansorge (2009), Metzker (2010), Shendure and Ji (2008) and van Dijk *et al.* (2014). Specifically, the DNA fragments that have been ligated to oligo adapters are released into an oil emulsion where they bind to complementary sequences on the surface of the beads. The beads with the bound DNA fragments together with reagents for PCR amplification are individually captured in oil emulsion capsules that are subjected to PCR temperature cycling to amplify each DNA template. After emulsion PCR, the emulsion capsules are broken and the beads are treated with denaturant to wash away untethered nucleic acid leaving the beads for a hybridization-based enrichment. Primers for sequencing are then added which hybridise to the adapter at a specific position and orientation. Pre-incubation of the beads with polymerase and single-stranded binding protein is followed by these being deposited into a 454 microfabricated array of picoliter wells, with only one bead per well. Added to the well are immobilized enzymes such as ATP sulfurylase and luciferase that are required for pyrosequencing. A

single species of labelled nucleotide is added at a time to each well and if incorporation is successful, the pyrophosphate is then released which generate a light signal that is monitored live by a fiber-optic bundle. Roche 454 is capable of generating large reads with a maximum of ~1000 nts in a short period of time. However, the limitation of this platform is that the reagents are costly and have high insertion and deletion error rates in homopolymers repeats of identical bases (AAA or CCC) (Ansorge, 2009; Metzker, 2010; Shendure & Ji, 2008).

1.6.2 Illumina platform

Illumina system employs clonal or bridge amplification for DNA template preparation and sequencing by synthesis (SBS), as reviewed in Anandhakumar *et al.* (2015), Ansorge (2009), Buermans and den Dunnen (2014), Reuter *et al.* (2015) and Shendure and Ji (2008). The DNA library consists of about 300 bp with each end having ligated oligo adapters. The DNA fragments are initially denatured from dsDNA into individual ssDNA molecules before being loaded onto the flow-cells. Once entering the flow-cell, one end of the oligo adapters is fixed to a complementary adapter on the inside surface of the flow cells. A polymerase creates a complement of the hybridized fragment then the double-stranded (ds) molecule is denatured and the original strand is washed away. Bridge amplification is then proceeded via isothermal amplification process, where each ssDNA fragment that is fixed at one end to the surface creates a bridge by hybridising the other free end to the complementary adapter on the surface forming a ds bridge. The ds bridge is then denatured using formamide forming two ssDNA molecules that are attached to the surface of the flow cell and the process is repeated. The adapters on the surface of the support are acting as forward and reverse primers for PCR amplification and after a few cycles, clusters of approximately 1000 copies of ssDNA fragments are created at the surface of the flow-cells. The SBS then occur with the addition of a reaction mixture that contains primers, reversible terminator nucleotides for the four bases each fluorescently labelled with a different dye and a DNA polymerase. After incorporation into the ssDNA strand by the DNA polymerase, two chemical bonds are removed from the reversible terminator nucleotide causing the fluorescent label to detach or cleave off from the nucleotide base. The releasing of the fluorophore from the nucleotide base causes a burst of light that is captured specifically to each four bases. Illumina is currently the most widely used NGS platform in metagenomics studies and it is the NGS platform that was used in this study. Illumina sequencing platform is cost effective and has a low error rate with an average raw error rates of 1-1.5% and high raw base accuracy of

>99.5% (Shendure & Ji, 2008). Depending on the Illumina model, the resulting single pair-end reads range from 2 x 100 bp to 2 x 300 bp. These read lengths, however, are much smaller compared to the Roche 454 platform (Buermans & den Dunnen, 2014; Metzker, 2010; Shendure & Ji, 2008).

1.6.3 SOLiD platform

In the SOLiD system, template DNA fragments, primers and PCR reaction components are amplified on microbeads by oil emulsion PCR to obtain clonal populations of the DNA templates. The recovered beads with the attached clonally amplified DNA fragments are then fixed covalently to the surface of a glass slide. Universal primers are then added which are complementary to the anchor adapter sequence. Following the annealing of the primer, a set of unique fluorophore-tagged probes is added. The fluorescently tagged probes contain 16 possible di-nucleotides sequences with combinations of two bases and first two positions are complementary to the recognition site. DNA ligase ligates the probe that is complementary to the universal primer. Following ligation, fluorescence images of all the DNA templates are captured and cleavage of the dye with silver ions follows leaving a reactive 5' phosphate group and the cycle is repeated, see (Anandhakumar *et al.*, 2015; Shendure & Ji, 2008; van Dijk *et al.*, 2014) for review. Note that SOLiD system sequences each DNA template twice in two separate cycles to provide high error correction rate and quality sequencing (Abbasian *et al.*, 2015; Anandhakumar *et al.*, 2015; Liu *et al.*, 2012; Metzker, 2010). However, the disadvantages of the SOLiD system include long run times and read length of up to 85 bp which is very short in comparison with other NSG platforms (Abbasian *et al.*, 2015; Liu *et al.*, 2012; Metzker, 2010; Shendure & Ji, 2008).

1.6.4 Ion torrent

Ion torrent sequencing platform is quite similar to Roche 454 except that it contains hydrogen ions (H^+) that are released when two nucleotides are incorporated which it is detected using an ion sensor, as reviewed in Buermans and den Dunnen (2014), Anandhakumar *et al.* (2015) and Reuter *et al.* (2015). Specifically, ion torrent contains a semiconductor chip that has millions of microwells that can decode the DNA sequence by measuring pH changes that are induced by the release of H^+ upon incorporation of nucleotides onto the new complementary DNA strand (Rothberg *et al.*, 2011). The pH changes are detected by an ion sensor at the bottom of the microwell which is then converted into a voltage signal that works accordingly

to the type and number of nucleotide incorporated (Rothberg *et al.*, 2011). DNA sequencing in ion torrent starts with a sample of DNA that is fragmented followed by each fragment attaching to its own bead and copied until it covers the whole bead. The beads with the attached DNA fragments are incubated in microwells (one bead per well) together with DNA polymerase and two type of nucleotides. The DNA polymerase adds complementary nucleotide complementary to the template strand which then followed by releasing of a H^+ that activates the ion sensor which converts the voltage signal into base sequences. At the end of each cycle, the remaining nucleotides are washed off before the next cycle begins. Note that multiple nucleotides will be added in a single cycle if there are repeated bases in the template strand meaning that the voltage signal will be stronger depending on the number of H^+ released. Ion torrent is affordable and the read length can be up to 200-400 bp. The run time is 4-8 hrs making sequencing faster compared to other sequencing platforms (Anandhakumar *et al.*, 2015; Buermans & den Dunnen, 2014; Liu *et al.*, 2012). However, insertions and deletions errors usually occur and also high error rates of homopolymer repeats due to addition of multiple nucleotides (Anandhakumar *et al.*, 2015; Liu *et al.*, 2012; Rothberg *et al.*, 2011).

1.6.5 Nanopore technology

The most significant progress seen in the evolution of NGS platforms is the emerging single-molecule sequencing approach known as nanopore technology, as reviewed in Bayley (2015) and Reuter *et al.* (2015). The principle of nanopore sequencing is based on the transition of DNA or single nucleotides through a small channel, with each sequencing flow well consists of many independent micro-wells (Wang *et al.*, 2014). Each micro-well is surrounded by a synthetic bilayer of nanopores. The need for a library preparation which is a crucial step in other NGS platforms is not required in nanopore technology, with sequencing can be carried out with or without shearing of DNA. However, adapters are required, with one adapter binding to a motor enzyme and a molecular tether and the other adapter consist of a hairpin oligonucleotide that is bound to a second motor protein. During sequencing, the changes in the induced electrical current are measured as a molecular motor protein passes the DNA strand through the nanopores. The first marketed nanopore technology device released by Oxford Nanopore Technologies is the so called MinION (Loman & Watson, 2015). MinION is relatively small and works off USB- port. The nanopore technology is portable, and is able to generate large read length relatively fast (Bayley, 2015; Reuter *et al.*, 2015), with a single

run generating approximately 16,000 total reads in 18 hr with a maximum read lengths of >60 kb (Ashton *et al.*, 2015). Unfortunately the error rates of the MinION nanopore sequencer are quite high and also high run failure rate; with substitution having error rates of 5.1%, deletion rates of 7.8% and 4.9% error rates of insertion (Jain *et al.*, 2015). The main approach that is being championed to deal with these high error rates is resequencing of the DNA using Nanopore sequencing data in conjunction with Illumina data.

1.6.6 PacBio platform

Pacific Biosciences (PacBio) platform is another leading NGS platform using a molecule sequencing approach (Buermans & den Dunnen, 2014; Niedringhaus *et al.*, 2011; Reuter *et al.*, 2015). It works by directly measuring DNA polymerase that incorporated fluorescent labelled nucleotides onto a complementary sequencing template. This sequencing platform contains highly parallel zero-mode waveguide (ZMW) nanostructures arrays that are packed onto a surface of specialised foundation that maintain optical confocal (Niedringhaus *et al.*, 2011). The DNA polymerases are loaded to the bottom of each ZMW where they can process the four phospholinked nucleotides that are fluorescently labelled (Niedringhaus *et al.*, 2011). The PacBio instrument can consecutively sequence both the sense and antisense strand of a dsDNA fragment by ligating hairpin loops to the ends of the fragments. The main advantages of using this technology for sequencing is that the amplification of sequencing fragments is not required, the time taken for sequencing is short and the ability to sequence long reads of up to 3000 bases (Coupland *et al.*, 2012; Niedringhaus *et al.*, 2011). Moreover, it can used for direct sequencing of small DNA molecules without standard library preparation implying that 1 ng of DNA is sufficient for generating reasonable sequence data (Coupland *et al.*, 2012). However, limitations include the inefficient loading of DNA polymerase in ZMWs, polymerase degrade in ZMWs, low single-pass accuracy of sequencing and instrument is relatively expensive (Niedringhaus *et al.*, 2011).

1.7 Discovery of novel CRESS DNA viruses

Viral metagenomic approaches using NGS have led to the discovery of novel circular ssDNA viral genomes that share some similarity to the known families *Geminiviridae*, *Nanoviridae*, *Circoviridae*, and the recently proposed family Genomoviridae. Before viral metagenomic studies became technologically feasible in the last decade, only a few studies were able to identify CRESS DNA viruses. Since then, with cost effective metagenomics sequencing

platforms, there has been an explosion of interest in CRESS DNA viruses. Metagenomic studies have enabled the identification of novel CRESS DNA viruses in various environmental samples including faecal matter, air, invertebrates, sewage, soil and water. Environmental sampling enables the exploration of greater viral sequence space, revealing a complex array of viruses that are interacting with the environment. Metagenomic studies have shown that the viral diversity of CRESS DNA viruses is grossly underestimated.

1.7.1 CRESS DNA viruses identified from environmental samples

1.7.1.1 Faecal matter

Sampling faecal matter of animals not only allows for the identification of viruses that are associated with the animal itself and their diet but also those infecting organisms such as bacteria, fungi and invertebrates that are present in the faecal matters at time of sampling (Delwart & Li, 2012). Faecal sampling is also non-invasive to the animal hence it can be easily used for viral surveillance in ecosystems. Novel CRESS DNA viruses have been identified in various metagenomics studies of faecal matter sampled from humans (Castrignano *et al.*, 2013; Garigliany *et al.*, 2014; Li *et al.*, 2010a; Ng *et al.*, 2015; Phan *et al.*, 2015; Victoria *et al.*, 2009) and various animals (Blinkova *et al.*, 2010; Cheung *et al.*, 2015; Cheung *et al.*, 2014a; b; Cheung *et al.*, 2013; Garigliany *et al.*, 2014; Ge *et al.*, 2011; Ge *et al.*, 2012; Hanna *et al.*, 2015; Hansen *et al.*, 2015; Kim *et al.*, 2012; Li *et al.*, 2010a; Li *et al.*, 2010b; Li *et al.*, 2015; Lima *et al.*, 2015; Ng *et al.*, 2014; Phan *et al.*, 2011; Reuter *et al.*, 2014; Sachsenroder *et al.*, 2014; Sachsenroder *et al.*, 2012; Sasaki *et al.*, 2015; Shan *et al.*, 2011; Sikorski *et al.*, 2013; van den Brand *et al.*, 2012; van Doorn *et al.*, 2013; Woo *et al.*, 2014; Zhang *et al.*, 2014). In particular, novel CRESS DNA viruses have been identified in faeces of various bat species from a number of countries (Ge *et al.*, 2011; Ge *et al.*, 2012; Li *et al.*, 2010a; Lima *et al.*, 2015). Whether the CRESS DNA viruses recovered from bat faeces contains viruses shed by the infected bat or those that are associated with its diet and the environment, remains unknown.

1.7.1.2 Air

Airborne viruses that cause diseases have continued to be a key research focus however, the diversity of CRESS DNA viruses circulating in the air in general still remains unknown. To date, the only metagenomics study that have successfully identified CRESS DNA viruses in air samples, was collected from three different types of land use in Korea, a residential

district, a forest and an industrial complex (Whon *et al.*, 2012). Using a Roche 454 platform, the resulting sequences identified were mostly CRESS DNA viruses with geminivirus-like and gemycircularvirus-like sequences being the most abundant (Whon *et al.*, 2012). The main challenges of carrying out metagenomic studies of airborne viruses is the low concentration of viral particles in the air and lack of standardised air sampling protocols (Behzad *et al.*, 2015; Womack *et al.*, 2010).

1.7.1.3 Aquatic environments

CRESS DNA viruses have been identified in various aquatic environments, both marine and fresh water. A study that sampled the coastal waters of British Columbia, the Gulf of Mexico and Saanich Inlet identified a large number of highly divergent marine CRESS DNA viruses (Labonte & Suttle, 2013). CRESS DNA viruses have been identified from Antarctic, Arctic freshwater and other freshwater lakes (de Cárcer *et al.*, 2015; Lopez-Bueno *et al.*, 2009; Roux *et al.*, 2012; Zawar-Reza *et al.*, 2014; Zhong *et al.*, 2015). Moreover, CRESS DNA viruses have also been identified in reclaimed water, ballast water, rainwater, lagoon wastewater and perennial ponds (Alhamlan *et al.*, 2013; Fancello *et al.*, 2013; Kim *et al.*, 2015; Rosario *et al.*, 2009; Whon *et al.*, 2012).

1.7.1.4 Invertebrates

Eukaryotic CRESS DNA viruses initially known to only infect plants and animals but since 2011, various metagenomics studies have also identified these viruses in numerous terrestrial and aquatic invertebrates. CRESS DNA viruses have been identified to be associated with odonata (both adults and larvae) which are top-end insect predators and thus may accumulate CRESS DNA viruses from insect prey, (Dayaram *et al.*, 2014; Dayaram *et al.*, 2013b; Dayaram *et al.*, 2015b; Rosario *et al.*, 2012a; Rosario *et al.*, 2011; Rosario *et al.*, 2013). Various CRESS DNA viruses have also being identified in other insects including mosquitoes, whiteflies, ticks and cockroaches (Garigliany *et al.*, 2015; Ng *et al.*, 2011a; Ng *et al.*, 2011c; Padilla-Rodriguez *et al.*, 2013; Xia *et al.*, 2015). Recent studies have explored the diversity of CRESS DNA viruses in a variety of aquatic invertebrates, predominantly crustaceans (Rosario *et al.*, 2015) and also in *Daphnia spp*., marine copepods, Florida estuarine mollusc species, shrimp and Forbes sea star (Dayaram *et al.*, 2015a; Dayaram *et al.*, 2013a; Dunlap *et al.*, 2013; Fahsbender *et al.*, 2015; Hewson *et al.*, 2013; Ng *et al.*, 2013).

1.7.1.5 Sewage

Sampling both treated and untreated sewage systems is an ideal way of studying the diversity of viruses such as those that infect the environmental microbes which are associated with faecal matters and those that infect humans (Blinkova *et al.*, 2009; Cantalupo *et al.*, 2011; Ng *et al.*, 2012; Parsley *et al.*, 2010; Phan *et al.*, 2015; Symonds *et al.*, 2009; Tamaki *et al.*, 2012). CRESS DNA viruses have been identified in untreated sewage samples and sewage oxidation pond (Blinkova *et al.*, 2009; Cantalupo *et al.*, 2011; Kraberger *et al.*, 2015; Ng *et al.*, 2012; Phan *et al.*, 2015).

1.7.1.6 Soil and sediments

The knowledge on the viruses circulating in soil is rather limited. A study on soil from rice paddy in Korea (Kim *et al.*, 2008) identified novel CRESS DNA viruses and recently reported that agricultural soil contain diverse CRESS DNA viruses (Reavy *et al.*, 2015). Sequence reads obtaining from deep-sea sediments were related to some members of CRESS DNA viruses but distinct from those in marines and freshwaters environments (Yoshida *et al.*, 2013).

1.8 Aims and objectives of this thesis research

1.8.1 Knowledge of viruses circulating in the Tongan archipelago

Previous studies on viruses in Tonga have been biased toward agricultural pathogens of economically important crops (Davis *et al.*, 2006; Davis & Ruabete, 2010; Kenyon *et al.*, 2008; Pearson & Pone, 1988; Stainton *et al.*, 2015a; Stainton *et al.*, 2012; Stainton *et al.*, 2015b). Others have mainly focused on viruses associated with human health (Chen & Maguire, 1990; Gubler *et al.*, 1978; Han *et al.*, 2014; Nelson *et al.*, 2014; Ushijima *et al.*, 1990; Wainwright *et al.*, 1986). Very little is known about the diversity of CRESS DNA viruses circulating in Tonga and to date only a handful of these have been identified. These include a nanovirus (*Banana bunchy top virus*) (Rosario *et al.*, 2012a; Stainton *et al.*, 2015a; Stainton *et al.*, 2016) which infects bananas, a cyclovirus (Dragonfly cyclovirus - 1) recovered from *Diplacodes bipunctata*, *Pantala flavescens*, *Tholymis tillarga* (Rosario & Breitbart, 2011; Rosario *et al.*, 2012a), two gemycircularviruses (Poaceae-associated gemycircularvirus - 1 recovered from *Brachiaria deflexa* and *Saccharum hybrid* (Male *et al.*, 2015; Rosario *et al.*, 2012a) and two unclassified CRESS DNA viruses (Dragonfly

circularisvirus from *P. flavescens* and Dragonfly orbiculatusvirus from *D. bipunctata*) (Rosario *et al.*, 2012a).

As part of this master thesis research, a viral metagenomics study was undertaken on the faecal matter of Pacific flying fox (*Pteropus tonganus*) with an aim to identify novel CRESS DNA viruses. Sampling animal faecal matter is non-invasive and previous studies have demonstrated that sampling animal faecal matter, including those of bats, contains a high diversity of viruses. To date, nothing is known about viruses associated with bats in the South Pacific. Bats, mammals of the order Chiroptera, in general are recognized as natural hosts and perhaps reservoirs of a large variety of both RNA and DNA viruses (~200 viruses of 27 families) (Calisher *et al.*, 2006), some of these viruses are responsible for emerging infectious diseases (Moratelli & Calisher, 2015; Smith & Wang, 2013). Zoonotic viruses like Ebola, Marburg, Nipah, Hendra, Rabies and coronaviruses have all been detected in bats (Brook & Dobson, 2015; Calisher *et al.*, 2006; Han *et al.*, 2015; Omatsu *et al.*, 2007; Plowright *et al.*, 2015; Wong *et al.*, 2007).

1.8.2 Bats as reservoir hosts of viruses

It has been speculated that bats have a high capacity for acting as reservoir hosts for zoonotic diseases (Brook & Dobson, 2015; Calisher *et al.*, 2006; Dobson, 2005). Reasons for this include bat specific tendencies/ecology such as, the social nature of bats with many individuals roosting together, long lifespans (with one species of bat living up to 35years), and potentially early co-speciation of bats and some zoonotic viruses as both have ancient evolutionary origins as reviewed in Calisher *et al.* (2006). Flight is also important in the spread of viruses, with many species able to fly long distances and potentially spread viruses between bat species (Calisher *et al.*, 2006). Flying also limits the amount of food bats can ingest at one time, therefore bats chew fruit to extract the high energy nutrients and discard the heavy partially digested fruit. These feeding habits of bats may also play a role in viral outbreaks (Dobson, 2005). The discarded fruits, which contain bat saliva, have been observed in the vicinity of a number of viral outbreaks (Chua *et al.*, 2002). Bats are often found in close proximity to humans and are also hunted and eaten in some regions as bush meat (Kamins *et al.*, 2011; Struebig *et al.*, 2007). Deforestation causes bat colonies to migrate closer to areas populated by humans in search of food and roosting sites as well as human mining in caves, increased the potential contact between bats and humans(Smith & Wang, 2013). Transmission events from bats to humans and other animals is generally via direct

contact through bites, handling or being eaten and indirect contacts in the form of aerosol, urine or faeces (Baker *et al.*, 2013; Han *et al.*, 2015; Smith & Wang, 2013; Wu *et al.*, 2015).

1.8.2.1 Pacific flying fox (*Pteropus tonganus*)

The bat species *Pteropus tonganus* (Family: Pteropodidae) also known as Pacific flying foxes of the genus Pteropus are frugivores found on small tropical islands (Pierson & Rainey, 1992). *P. tonganus* is the most widespread flying fox in the Pacific and are widely distributed throughout the South Pacific islands including Papua New Guinea, Solomon Islands, Vanuatu, New Caledonia, Fiji, Samoa, American Samoa, Niue, Cook Islands and Tonga (Pierson & Rainey, 1992). Few observations reported that flying foxes can fly between islands depending on the availability of food (McConkey & Drake, 2007; McConkey *et al.*, 2004; Pierson & Rainey, 1992; Rinke, 1991). There are many other species in the genus Pteropus, however, *P. tonganus* is the only species found in the Kingdom of Tonga (Miller & Wilson, 1997).

P. tonganus can live for ~30 years and they roost on trees as shown in Figure 1.5. They commonly feed on fruits, leaves and nectar for which they play a crucial role in pollination and seed dispersal (Banack, 1998; Nelson *et al.*, 2005; Pierson & Rainey, 1992). Pacific flying foxes were once considered agricultural pests because they feed on economically important fruits (Wiles & Payne, 1986). Interestingly, the damage caused to fruit crops by flying foxes resulted in the intentional introduction of avian cholera into Samoa to eradicate bats and various other birds (Spennemann & Gary, 2002).

In Tonga, Pacific flying foxes hold cultural values and are protected by the Tongan royal family from human exploitation but in other Pacific islands they can be freely hunted (Wiles & Fujita, 1992; Wiles & Payne, 1986). In some Pacific islands, Pacific flying foxes are considered an important traditional food and are commonly consumed (Wiles & Payne, 1986). However, although flying foxes are not a traditional food and are protected in Tonga, there are still incidences of bats being killed and eaten for bush meat which increases the risk of encountering diseases from bats. The majority of the roosting sites of *P.tonganus* in Tonga are located within the villages where they are protected and only few roosting sites are located in the push areas. Pacific flying foxes are found in groups with thousands of bats roosting together, with numbers likely higher in Tonga due to their protected status. This

gregarious lifestyle of Pacific flying foxes could potentially amplify viruses in natural settings.

Bats are natural hosts for a wide range of viruses and to date, no baseline analyses of the ssDNA viruses of Pacific flying foxes faeces in Tonga has been carried out. As Tonga is my home country, I was interested in the continuing effort of expanding the diversity of CRESS DNA viruses in Tonga.

A B

Figure 1.5: Two Pacific flying foxes roosting in a tree (A), part of a larger roosting site located at Lapaha (Takuilau) in Tongatapu, Tonga (B).

1.8.3 Specific aims of this study

Using a viral metagenomics approach, this study aimed to

1. Identify CRESS DNA viruses associated with *P. tonganus* faecal matter
2. Determine whether there are differences in CRESS DNA viral assemblages at four *P. tonganus* roosting sites on Tongatapu, the main island of Tonga
3. Determine whether any CRESS DNA viruses are persistently associated with *P. tonganus* faeces
4. Given that *P. tonganus* are frugivores, identify putative plant-infecting viruses circulating in Tonga via faecal sampling

1.9 References

Abbasian, F., Lockington, R., Megharaj, M. & Naidu, R. (2015). The integration of sequencing and bioinformatics in metagenomics. *Rev Environ Sci Bio* **14**, 357-383.

Abraham, A. D., Varrelmann, M. & Vetten, H. J. (2012). Three Distinct Nanoviruses, One of Which Represents a New Species, Infect Faba Bean in Ethiopia. *Plant Disease* **96**, 1045-1053.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M. & Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol Plant Pathol* **10**, 537-545.

Adams, M. J. & Carstens, E. B. (2012). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). *Arch Virol* **157**, 1411-1422.

Alhamlan, F. S., Ederer, M. M., Brown, C. J., Coats, E. R. & Crawford, R. L. (2013). Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *J Microbiol Methods* **92**, 183-188.

Allan, G. M. & Ellis, J. A. (2000). Porcine circoviruses: a review. *J Vet Diagn Invest* **12**, 3-14.

Allander, T. & Andersson, B. (2012). Human Bocavirus and methods of diagnosis and treatment: Google Patents.

Amin, I., Mansoor, S., Amrao, L., Hussain, M., Irum, S., Zafar, Y., Bull, S. E. & Briddon, R. W. (2006). Mobilisation into cotton and spread of a recombinant cotton leaf curl disease satellite. *Arch Virol* **151**, 2055-2065.

Anandhakumar, C., Kizaki, S., Bando, T., Pandian, G. N. & Sugiyama, H. (2015). Advancing small-molecule-based chemical biology with next-generation sequencing technologies. *Chembiochem* **16**, 20-38.

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *N Biotechnol* **25**, 195-203.

Ashton, P. M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J. & O'Grady, J. (2015). MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat Biotechnol* **33**, 296-300.

Baker, K. S., Leggett, R. M., Bexfield, N. H., Alston, M., Daly, G., Todd, S., Tachedjian, M., Holmes, C. E., Crameri, S., Wang, L. F., Heeney, J. L., Suu-Ire, R., Kellam, P., Cunningham, A. A., Wood, J. L., Caccamo, M. & Murcia, P. R. (2013). Metagenomic study of the viruses of African straw-coloured fruit bats: detection of a chiropteran poxvirus and isolation of a novel adenovirus. *Virology* **441**, 95-106.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological reviews* **35**, 235-241.

Banack, S. A. (1998). Diet selection and resource use by flying foxes (Genus Pteropus). *Ecology* **79**, 1949-1967.

Bayley, H. (2015). Nanopore sequencing: from imagination to reality. *Clin Chem* **61**, 25-31.

Behzad, H., Gojobori, T. & Mineta, K. (2015). Challenges and opportunities of airborne metagenomics. *Genome Biol Evol* **7**, 1216-1226.

Biagini, P., de Micco, P. & de Lamballerie, X. (2006). Identification of a third member of the Anellovirus genus ("small anellovirus") in French blood donors. *Arch Virol* **151**, 405-408.

Biagini, P., Uch, R., Belhouchet, M., Attoui, H., Cantaloube, J. F., Brisbarre, N. & de Micco, P. (2007). Circular genomes related to anelloviruses identified in human and animal samples by using a combined rolling-circle amplification/sequence-independent single primer amplification approach. *J Gen Virol* **88**, 2696-2701.

Blinkova, O., Rosario, K., Li, L., Kapoor, A., Slikas, B., Bernardin, F., Breitbart, M. & Delwart, E. (2009). Frequent detection of highly diverse variants of cardiovirus, cosavirus, bocavirus, and circovirus in sewage samples collected in the United States. *J Clin Microbiol* **47**, 3507-3513.

Blinkova, O., Victoria, J., Li, Y., Keele, B. F., Sanz, C., Ndjango, J. B., Peeters, M., Travis, D., Lonsdorf, E. V., Wilson, M. L., Pusey, A. E., Hahn, B. H. & Delwart, E. L. (2010). Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* **91**, 74-86.

Bouzari, M. & Salmanizadeh, S. (2015). Detection of Torque teno midi virus/Small anellovirus (TTMDV/SAV) in the sera of domestic village chickens and its vertical transmission from hen to eggs. *Iranian Journal of Veterinary Research* **16**, 110-113.

Breitbart, M. & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in microbiology* **13**, 278-284.

Brentlinger, K. L., Hafenstein, S., Novak, C. R., Fane, B. A., Borgon, R., McKenna, R. & Agbandje-McKenna, M. (2002). Microviridae, a family divided: isolation, characterization, and genome sequence of φMH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus. *Journal of bacteriology* **184**, 1089-1094.

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-ul-Rehman, M. S. & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evol Biol* **10**, 97.

Brook, C. E. & Dobson, A. P. (2015). Bats as 'special' reservoirs for emerging zoonotic pathogens. *Trends Microbiol* **23**, 172-180.

Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C., Fiallo-Olive, E., Briddon, R. W., Hernandez-Zepeda, C., Idris, A., Malathi, V. G., Martin, D. P., Rivera-Bustamante, R., Ueda, S. & Varsani, A. (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Arch Virol* **160**, 1593-1619.

Bryson, S. J., Thurber, A. R., Correa, A. M., Orphan, V. J. & Vega Thurber, R. (2015). A novel sister clade to the enterobacteria microviruses (family Microviridae) identified in methane seep sediments. *Environ Microbiol* **17**, 3708-3721.

Buermans, H. P. & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta* **1842**, 1932-1941.

Burian, Z., Szabo, H., Szekely, G., Gyurkovits, K., Pankovics, P., Farkas, T. & Reuter, G. (2011). Detection and follow-up of torque teno midi virus ("small anelloviruses") in nasopharyngeal aspirates and three other human body fluids in children. *Arch Virol* **156**, 1537-1541.

Burns, T. M., Harding, R. M. & Dale, J. L. (1995). The genome organization of banana bunchy top virus: analysis of six ssDNA components. *J Gen Virol* **76 (Pt 6)**, 1471-1482.

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. (2006). Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev* **19**, 531-545.

Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., Grabe, M., Hendrix, R. W., Girones, R., Wang, D. & Pipas, J. M. (2011). Raw sewage harbors diverse viral populations. *MBio* **2**, e00180-00111.

Castrignano, S. B., Nagasse-Sugahara, T. K., Kisielius, J. J., Ueda-Ito, M., Brandao, P. E. & Curti, S. P. (2013). Two novel circo-like viruses detected in human feces: complete genome sequencing and electron microscopy analysis. *Virus Res* **178**, 364-373.

Chen, L. F., Rojas, M., Kon, T., Gamby, K., Xoconostle-Cazares, B. & Gilbertson, R. L. (2009). A severe symptom phenotype in tomato in Mali is caused by a reassortant between a novel recombinant begomovirus (Tomato yellow leaf curl Mali virus) and a betasatellite. *Mol Plant Pathol* **10**, 415-430.

Chen, W. B. & Maguire, T. (1990). Nucleotide sequence of the envelope glycoprotein gene of a dengue-2 virus isolated during an epidemic of benign dengue fever in Tonga in 1974. *Nucleic Acids Res* **18**, 5889.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. & Pogranichniy, R. M. (2015). Identification of several clades of novel single-stranded circular DNA viruses with conserved stem-loop structures in pig feces. *Arch Virol* **160**, 353-358.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. L. & Pogranichniy, R. M. (2014a). Identification of a novel single-stranded circular DNA virus in pig feces. *Genome Announc* **2**, e00347-00314.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. L. & Pogranichniy, R. M. (2014b). Unique circovirus-like genome detected in pig feces. *Genome Announc* **2**, e00251-00214.

Cheung, A. K., Ng, T. F., Lager, K. M., Bayles, D. O., Alt, D. P., Delwart, E. L., Pogranichniy, R. M. & Kehrli, M. E., Jr. (2013). A divergent clade of circular single-stranded DNA viruses from pig feces. *Arch Virol* **158**, 2157-2162.

Chu, P. W. & Helms, K. (1988). Novel virus-like particles containing circular single-stranded DNAs associated with subterranean clover stunt disease. *Virology* **167**, 38-49.

Chua, K. B., Koh, C. L., Hooi, P. S., Wee, K. F., Khong, J. H., Chua, B. H., Chan, Y. P., Lim, M. E. & Lam, S. K. (2002). Isolation of Nipah virus from Malaysian Island flying-foxes. *Microbes Infect* **4**, 145-151.

Cibulski, S. P., Teixeira, T. F., de Sales Lima, F. E., do Santos, H. F., Franco, A. C. & Roehe, P. M. (2014). A Novel Anelloviridae Species Detected in Tadarida brasiliensis Bats: First Sequence of a Chiropteran Anellovirus. *Genome Announc* **2**, e01028-01014.

Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. A., Mukha, D. V., Pintel, D. J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D. & Davison, A. J. (2014). The family Parvoviridae. *Arch Virol* **159**, 1239-1247.

Coupland, P., Chandra, T., Quail, M., Reik, W. & Swerdlow, H. (2012). Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. *Biotechniques* **53**, 365-372.

Csagola, A., Kecskemeti, S., Kardos, G., Kiss, I. & Tuboly, T. (2006). Genetic characterization of type 2 porcine circoviruses detected in Hungarian wild boars. *Arch Virol* **151**, 495-507.

Davis, R., Tupouniua, S., Amice, R., Taufa, L. & Jones, P. (2006). *Surveys for Plant Diseases Caused by Viruses & Virus-like Pathogens in Tonga & New Caledonia*: SPC Land Resources Division.

Davis, R. I. & Ruabete, T. K. (2010). Records of plant pathogenic viruses and virus-like agents from 22 Pacific island countries and territories: a review and an update. *Australasian Plant Pathology* **39**, 265-291.

Dayaram, A., Galatowitsch, M., Harding, J. S., Arguello-Astorga, G. R. & Varsani, A. (2014). Novel circular DNA viruses identified in Procordulia grayi and Xanthocnemis zealandica larvae using metagenomic approaches. *Infect Genet Evol* **22**, 134-141.

Dayaram, A., Goldstien, S., Arguello-Astorga, G. R., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2015a). Diverse small circular DNA viruses circulating amongst estuarine molluscs. *Infect Genet Evol* **31**, 284-295.

Dayaram, A., Goldstien, S., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013a). Identification of starling circovirus in an estuarine mollusc (Amphibola crenata) in New Zealand using metagenomic approaches. *Genome announcements* **1**, e00278-00213.

Dayaram, A., Potter, K. A., Moline, A. B., Rosenstein, D. D., Marinov, M., Thomas, J. E., Breitbart, M., Rosario, K., Arguello-Astorga, G. R. & Varsani, A. (2013b). High global diversity of cycloviruses amongst dragonflies. *J Gen Virol* **94**, 1827-1840.

Dayaram, A., Potter, K. A., Pailes, R., Marinov, M., Rosenstein, D. D. & Varsani, A. (2015b). Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infect Genet Evol* **30**, 278-287.

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, M., Abdoul-Karime, A. L., Abdou-Chakour, C., Reynaud, B., Harkins, G. W., Varsani, A., Martin, D. P. & Lett, J. M. (2012). East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC Evol Biol* **12**, 228.

de Cárcer, D. A., López-Bueno, A., Pearce, D. A. & Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Science Advances* **1**, e1400127.

Delwart, E. & Li, L. (2012). Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other Rep encoding small circular ssDNA genomes. *Virus Res* **164**, 114-121.

Delwart, E. L. (2007). Viral metagenomics. *Rev Med Virol* **17**, 115-131.

Delwart, E. L. & Jones, M. S. (2014). Human parvovirus: Google Patents.

Dobson, A. P. (2005). What links bats to emerging infectious diseases? *Science* **310**, 628-629.

Domingo, E. (1994). Mutation rates and rapid evolution of RNA viruses. *The evolutionary biology of viruses*, 161-184.

Duffy, S. & Holmes, E. C. (2008). Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *J Virol* **82**, 957-965.

Duffy, S., Shackelton, L. A. & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* **9**, 267-276.

Dunlap, D. S., Ng, T. F., Rosario, K., Barbosa, J. G., Greco, A. M., Breitbart, M. & Hewson, I. (2013). Molecular and microscopic evidence of viruses in marine copepods. *Proc Natl Acad Sci U S A* **110**, 1375-1380.

Edwards, R. A. & Rohwer, F. (2005). Viral metagenomics. *Nat Rev Microbiol* **3**, 504-510.

Fahsbender, E., Hewson, I., Rosario, K., Tuttle, A. D., Varsani, A. & Breitbart, M. (2015). Discovery of a novel circular DNA virus in the Forbes sea star, Asterias forbesi. *Arch Virol* **160**, 2349-2351.

Fancello, L., Trape, S., Robert, C., Boyer, M., Popgeorgiev, N., Raoult, D. & Desnues, C. (2013). Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. *ISME J* **7**, 359-369.

Fauquet, C., Bisaro, D., Briddon, R., Brown, J., Harrison, B., Rybicki, E., Stenger, D. & Stanley, J. (2003). Virology division news: revision of taxonomic criteria for species

demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of virology* **148**, 405-421.

Firth, C., Charleston, M. A., Duffy, S., Shapiro, B. & Holmes, E. C. (2009). Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. *J Virol* **83**, 12813-12821.

Forterre, P. & Prangishvili, D. (2009). The Great Billion-year War between Ribosome-and Capsid-encoding Organisms (Cells and Viruses) as the Major Source of Evolutionary Novelties. *Annals of the New York Academy of Sciences* **1178**, 65-77.

Garigliany, M.-M., Hagen, R. M., Frickmann, H., May, J., Schwarz, N. G., Perse, A., Jöst, H., Börstler, J., Shahhosseini, N. & Desmecht, D. (2014). Cyclovirus CyCV-VN species distribution is not limited to Vietnam and extends to Africa. *Scientific reports* **4**.

Garigliany, M. M., Borstler, J., Jost, H., Badusche, M., Desmecht, D., Schmidt-Chanasit, J. & Cadar, D. (2015). Characterization of a novel circo-like virus in Aedes vexans mosquitoes from Germany: evidence for a new genus within the family Circoviridae. *J Gen Virol* **96**, 915-920.

Ge, L., Zhang, J., Zhou, X. & Li, H. (2007). Genetic structure and population variability of tomato yellow leaf curl China virus. *J Virol* **81**, 5902-5907.

Ge, X., Li, J., Peng, C., Wu, L., Yang, X., Wu, Y., Zhang, Y. & Shi, Z. (2011). Genetic diversity of novel circular ssDNA viruses in bats in China. *J Gen Virol* **92**, 2646-2653.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y. & Shi, Z. (2012). Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* **86**, 4620-4630.

Grigoras, I., Ginzo, A. I., Martin, D. P., Varsani, A., Romero, J., Mammadov, A., Huseynova, I. M., Aliyev, J. A., Kheyr-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H. J. & Gronenborn, B. (2014). Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *J Gen Virol* **95**, 1178-1191.

Grigoras, I., Gronenborn, B. & Vetten, H. J. (2010a). First Report of a Nanovirus Disease of Pea in Germany. *Plant Disease* **94**, 642-643.

Grigoras, I., Timchenko, T., Grande-Perez, A., Katul, L., Vetten, H. J. & Gronenborn, B. (2010b). High variability and rapid evolution of a nanovirus. *J Virol* **84**, 9105-9117.

Grigoras, I., Timchenko, T., Katul, L., Grande-Perez, A., Vetten, H. J. & Gronenborn, B. (2009). Reconstitution of authentic nanovirus from multiple cloned DNAs. *J Virol* **83**, 10778-10787.

Gubler, D. J., Reed, D., Rosen, L. & Hitchcock, J. R., Jr. (1978). Epidemiologic, clinical, and virologic observations on dengue in the Kingdom of Tonga. *Am J Trop Med Hyg* **27**, 581-589.

Gutierrez, C. (1999). Geminivirus DNA replication. *Cell Mol Life Sci* **56**, 313-329.

Hafner, G. J., Stafford, M. R., Wolter, L. C., Harding, R. M. & Dale, J. L. (1997). Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J Gen Virol* **78 (Pt 7)**, 1795-1799.

Han, H. J., Wen, H. L., Zhou, C. M., Chen, F. F., Luo, L. M., Liu, J. W. & Yu, X. J. (2015). Bats as reservoirs of severe emerging infectious diseases. *Virus Res* **205**, 1-6.

Han, M., Littlejohn, M., Yuen, L., Edwards, R., Devi, U., Bowden, S., Ning, Q., Locarnini, S. & Jackson, K. (2014). Molecular epidemiology of hepatitis delta virus in the Western Pacific region. *J Clin Virol* **61**, 34-39.

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S. & Robertson, D. (1999). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Sciences* **18**, 71-106.

Hanna, Z. R., Runckel, C., Fuchs, J., DeRisi, J. L., Mindell, D. P., Van Hemert, C., Handel, C. M. & Dumbacher, J. P. (2015). Isolation of a Complete Circular Virus Genome Sequence from an Alaskan Black-Capped Chickadee (Poecile atricapillus) Gastrointestinal Tract Sample. *Genome Announc* **3**, e01081-01015.

Hansen, T. A., Fridholm, H., Froslev, T. G., Kjartansdottir, K. R., Willerslev, E., Nielsen, L. P. & Hansen, A. J. (2015). New Type of Papillomavirus and Novel Circular Single Stranded DNA Virus Discovered in Urban Rattus norvegicus Using Circular DNA Enrichment and Metagenomics. *PLoS One* **10**, e0141952.

Harkins, G. W., Martin, D. P., Duffy, S., Monjane, A. L., Shepherd, D. N., Windram, O. P., Owor, B. E., Donaldson, L., van Antwerpen, T., Sayed, R. A., Flett, B., Ramusi, M., Rybicki, E. P., Peterschmitt, M. & Varsani, A. (2009). Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. *J Gen Virol* **90**, 3066-3074.

He, B., Li, Z., Yang, F., Zheng, J., Feng, Y., Guo, H., Li, Y., Wang, Y., Su, N., Zhang, F., Fan, Q. & Tu, C. (2013). Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel Mammalian viruses. *PLoS One* **8**, e61950.

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

Hewson, I., Ng, G., Li, W., LaBarre, B. A., Aguirre, I., Barbosa, J. G., Breitbart, M., Greco, A. W., Kearns, C. M. & Looi, A. (2013). Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a Daphnia-associated single-stranded DNA virus in two temperate lakes. *Limnology and Oceanography* **58**, 1605-1620.

Hu, J. M., Fu, H. C., Lin, C. H., Su, H. J. & Yeh, H. H. (2007). Reassortment and concerted evolution in banana bunchy top virus genomes. *J Virol* **81**, 1746-1761.

Hu, Z. Y., Li, G. H., Li, G. T., Yao, Q. & Chen, K. P. (2013). Bombyx mori bidensovirus: The type species of the new genus Bidensovirus in the new family Bidnaviridae. *Chinese Science Bulletin* **58**, 4528-4532.

Huang, Y. W., Ni, Y. Y., Dryman, B. A. & Meng, X. J. (2010). Multiple infection of porcine Torque teno virus in a single pig and characterization of the full-length genomic sequences of four U.S. prototype PTTV strains: implication for genotyping of PTTV. *Virology* **396**, 289-297.

Hughes, A. L. (2004). Birth-and-death evolution of protein-coding regions and concerted evolution of non-coding regions in the multi-component genomes of nanoviruses. *Mol Phylogenet Evol* **30**, 287-294.

Idris, A. M. & Brown, J. K. (2004). Cotton leaf crumple virus Is a Distinct Western Hemisphere Begomovirus Species with Complex Evolutionary Relationships Indicative of Recombination and Reassortment. *Phytopathology* **94**, 1068-1074.

Islam, M. N., Naqvi, A. R., Jan, A. T. & Haq, Q. M. R. (2010). Genetic diversity and possible evidence of recombination among Banana Bunchy Top Virus (BBTV) isolates. *Int Res J Microbiol* **1**, 1-12.

Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B. & Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nat Methods* **12**, 351-356.

Jiang, D., Fu, Y., Guoqing, L. & Ghabrial, S. A. (2013). Viruses of the plant pathogenic fungus Sclerotinia sclerotiorum. *Adv Virus Res* **86**, 215-248.

Julian, L., Piasecki, T., Chrzastek, K., Walters, M., Muhire, B., Harkins, G. W., Martin, D. P. & Varsani, A. (2013). Extensive recombination detected among beak and feather disease virus isolates from breeding facilities in Poland. *J Gen Virol* **94**, 1086-1095.

Kamins, A. O., Restif, O., Ntiamoa-Baidu, Y., Suu-Ire, R., Hayman, D. T., Cunningham, A. A., Wood, J. L. & Rowcliffe, J. M. (2011). Uncovering the fruit bat bushmeat commodity chain and the true extent of fruit bat hunting in Ghana, West Africa. *Biol Conserv* **144**, 3000-3008.

Kapoor, A., Dubovi, E. J., Henriquez-Rivera, J. A. & Lipkin, W. I. (2012). Complete genome sequence of the first canine circovirus. *J Virol* **86**, 7018.

Katul, L., Timchenko, T., Gronenborn, B. & Vetten, H. J. (1998). Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *Journal of General Virology* **79**, 3101-3109.

Kenyon, L., Lebas, B. S. & Seal, S. E. (2008). Yams (Dioscorea spp.) from the South Pacific Islands contain many novel badnaviruses: implications for international movement of yam germplasm. *Arch Virol* **153**, 877-889.

Kim, H. K., Park, S. J., Nguyen, V. G., Song, D. S., Moon, H. J., Kang, B. K. & Park, B. K. (2012). Identification of a novel single-stranded, circular DNA virus from bovine stool. *J Gen Virol* **93**, 635-639.

Kim, K. H., Chang, H. W., Nam, Y. D., Roh, S. W., Kim, M. S., Sung, Y., Jeon, C. O., Oh, H. M. & Bae, J. W. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl Environ Microbiol* **74**, 5975-5985.

Kim, Y., Aw, T. G., Teal, T. K. & Rose, J. B. (2015). Metagenomic Investigation of Viral Communities in Ballast Water. *Environ Sci Technol* **49**, 8396-8407.

King, A. M., Adams, M. J. & Lefkowitz, E. J. (2012). *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*: Elsevier.

Kraberger, S., Arguello-Astorga, G. R., Greenfield, L. G., Galilee, C., Law, D., Martin, D. P. & Varsani, A. (2015). Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. *Infect Genet Evol* **31**, 73-86.

Kraberger, S., Harkins, G. W., Kumari, S. G., Thomas, J. E., Schwinghamer, M. W., Sharman, M., Collings, D. A., Briddon, R. W., Martin, D. P. & Varsani, A. (2013). Evidence that dicot-infecting mastreviruses are particularly prone to inter-species recombination and have likely been circulating in Australia for longer than in Africa and the Middle East. *Virology* **444**, 282-291.

Kumari, S. G., Rodoni, B., Vetten, H. J., Loh, M. H., Freeman, A., Van Leur, J., Bao, S. Y. & Wang, X. M. (2010). Detection and Partial Characterization of Milk vetch dwarf virus Isolates from Faba Bean (Vicia faba L.) in Yunnan Province, China. *Journal of Phytopathology* **158**, 35-39.

Labonte, J. M. & Suttle, C. A. (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* **7**, 2169-2177.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J. & Gronenborn, B. (1995). In-Vitro Cleavage and Joining at the Viral Origin of Replication by the Replication Initiator Protein of Tomato Yellow Leaf Curl Virus. *P Natl Acad Sci USA* **92**, 3879-3883.

Lefeuvre, P., Lett, J. M., Varsani, A. & Martin, D. P. (2009). Widely conserved recombination patterns among single-stranded DNA viruses. *J Virol* **83**, 2697-2707.

Li, L., Kapoor, A., Slikas, B., Bamidele, O. S., Wang, C., Shaukat, S., Masroor, M. A., Wilson, M. L., Ndjango, J. B., Peeters, M., Gross-Camp, N. D., Muller, M. N., Hahn, B. H., Wolfe, N. D., Triki, H., Bartkus, J., Zaidi, S. Z. & Delwart, E. (2010a). Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol* **84**, 1674-1682.

Li, L., Shan, T., Soji, O. B., Alam, M. M., Kunz, T. H., Zaidi, S. Z. & Delwart, E. (2011). Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *J Gen Virol* **92**, 768-772.

Li, L., Victoria, J. G., Wang, C., Jones, M., Fellers, G. M., Kunz, T. H. & Delwart, E. (2010b). Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol* **84**, 6955-6965.

Li, W., Gu, Y., Shen, Q., Yang, S., Wang, X., Wan, Y. & Zhang, W. (2015). A novel gemycircularvirus from experimental rats. *Virus Genes* **51**, 302-305.

Lima, F. E., Cibulski, S. P., Dos Santos, H. F., Teixeira, T. F., Varela, A. P., Roehe, P. M., Delwart, E. & Franco, A. C. (2015). Genomic characterization of novel circular ssDNA viruses from insectivorous bats in Southern Brazil. *PLoS One* **10**, e0118070.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* **2012**, 251364.

Loman, N. J. & Watson, M. (2015). Successful test launch for nanopore sequencing. *Nat Methods* **12**, 303-304.

Lopez-Bueno, A., Tamames, J., Velazquez, D., Moya, A., Quesada, A. & Alcami, A. (2009). High diversity of the viral community from an Antarctic lake. *Science* **326**, 858-861.

Lorincz, M., Csagola, A., Farkas, S. L., Szekely, C. & Tuboly, T. (2011). First detection and analysis of a fish circovirus. *J Gen Virol* **92**, 1817-1821.

Male, M. F., Kami, V., Kraberger, S. & Varsani, A. (2015). Genome Sequences of Poaceae-Associated Gemycircularviruses from the Pacific Ocean Island of Tonga. *Genome Announc* **3**, e01144-01115.

Mandal, B. (2010). Advances in Small Isometric Multicomponent ssDNA Viruses Infecting Plants. *Indian J Virol* **21**, 18-30.

Mandal, B., Shilpi, S., Barman, A. R., Mandal, S. & Varma, A. (2013). Nine novel DNA components associated with the foorkey disease of large cardamom: evidence of a distinct babuvirus species in Nanoviridae. *Virus Res* **178**, 297-305.

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

Martin, D. P., Briddon, R. W. & Varsani, A. (2011b). Recombination patterns in dicot-infecting mastreviruses mirror those found in monocot-infecting mastreviruses. *Arch Virol* **156**, 1463-1469.

McConkey, K. R. & Drake, D. R. (2007). Indirect evidence that flying foxes track food resources among islands in a Pacific Archipelago. *Biotropica* **39**, 436-440.

McConkey, K. R., Drake, D. R., Franklin, J. & Tonga, F. (2004). Effects of Cyclone Waka on flying foxes (Pteropus tonganus) in the Vava'u Islands of Tonga. *Journal of Tropical Ecology* **20**, 555-561.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46.

Miller, C. A. & Wilson, D. E. (1997). Pteropus tonganus. *Mammalian Species*, 1-6.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., Lewis, J. D. & Bushman, F. D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* **21**, 1616-1625.

Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., Sako, Y., Forterre, P. & Prangishvili, D. (2012). Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proc Natl Acad Sci U S A* **109**, 13386-13391.

Mochizuki, T., Yoshida, T., Tanaka, R., Forterre, P., Sako, Y. & Prangishvili, D. (2010). Diversity of viruses of the hyperthermophilic archaeal genus Aeropyrum, and isolation of the Aeropyrum pernix bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. *Virology* **402**, 347-354.

Moratelli, R. & Calisher, C. H. (2015). Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Memórias do Instituto Oswaldo Cruz* **110**, 1-22.

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**, e108277.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Ibanez, J. T. & Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J Virol* **85**, 1182-1192.

Nawaz-Ul-Rehman, M. S., Nahid, N., Mansoor, S., Briddon, R. W. & Fauquet, C. M. (2010). Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* **405**, 300-308.

Nelson, J., Waggoner, J. J., Sahoo, M. K., Grant, P. M. & Pinsky, B. A. (2014). Encephalitis caused by Chikungunya virus in a traveler from the Kingdom of Tonga. *J Clin Microbiol* **52**, 3459-3461.

Nelson, S. L., Kunz, T. H. & Humphrey, S. R. (2005). Folivory in fruit bats: Leaves provide a natural source of calicum. *Journal of Chemical Ecology* **31**, 1683-1691.

Ng, T. F., Alavandi, S., Varsani, A., Burghart, S. & Breitbart, M. (2013). Metagenomic identification of a nodavirus and a circular ssDNA virus in semi-purified viral nucleic acids from the hepatopancreas of healthy Farfantepenaeus duorarum shrimp. *Dis Aquat Organ* **105**, 237-242.

Ng, T. F., Chen, L. F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P. D., Varsani, A., Kondov, N. O., Wong, W., Deng, X., Andrews, T. D., Moorman, B. J., Meulendyk, T., MacKay, G., Gilbertson, R. L. & Delwart, E. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. *Proc Natl Acad Sci U S A* **111**, 16842-16847.

Ng, T. F., Duffy, S., Polston, J. E., Bixby, E., Vallad, G. E. & Breitbart, M. (2011a). Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS One* **6**, e19050.

Ng, T. F., Manire, C., Borrowman, K., Langer, T., Ehrhart, L. & Breitbart, M. (2009). Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol* **83**, 2500-2509.

Ng, T. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B. S., Wommack, K. E. & Delwart, E. (2012). High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol* **86**, 12161-12175.

Ng, T. F., Wheeler, E., Greig, D., Waltzek, T. B., Gulland, F. & Breitbart, M. (2011b). Metagenomic identification of a novel anellovirus in Pacific harbor seal (Phoca vitulina richardsii) lung samples and its detection in samples from multiple years. *J Gen Virol* **92**, 1318-1323.

Ng, T. F., Willner, D. L., Lim, Y. W., Schmieder, R., Chau, B., Nilsson, C., Anthony, S., Ruan, Y., Rohwer, F. & Breitbart, M. (2011c). Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes. *PLoS One* **6**, e20579.

Ng, T. F. F., Zhang, W., Sachsenröder, J., Kondov, N. O., da Costa, A. C., Vega, E., Holtz, L. R., Wu, G., Wang, D. & Stine, C. O. (2015). A diverse group of small circular ssDNA viral genomes in human and non-human primate stools. *Virus Evolution* **1**, vev017.

Niedringhaus, T. P., Milanova, D., Kerby, M. B., Snyder, M. P. & Barron, A. E. (2011). Landscape of next-generation sequencing technologies. *Anal Chem* **83**, 4327-4341.

Ninomiya, M., Takahashi, M., Hoshino, Y., Ichiyama, K., Simmonds, P. & Okamoto, H. (2009). Analysis of the entire genomes of torque teno midi virus variants in chimpanzees: infrequent cross-species infection between humans and chimpanzees. *J Gen Virol* **90**, 347-358.

Nishiyama, S., Dutia, B. M., Stewart, J. P., Meredith, A. L., Shaw, D. J., Simmonds, P. & Sharp, C. P. (2014). Identification of novel anelloviruses with broad diversity in UK rodents. *J Gen Virol* **95**, 1544-1553.

Okamoto, H. (2009). History of discoveries and pathogenicity of TT viruses. In *Tt Viruses*, pp. 1-20: Springer.

Omatsu, T., Watanabe, S., Akashi, H. & Yoshikawa, Y. (2007). Biological characters of bats in relation to natural reservoir of emerging viruses. *Comp Immunol Microbiol Infect Dis* **30**, 357-374.

Owor, B. E., Martin, D. P., Shepherd, D. N., Edema, R., Monjane, A. L., Rybicki, E. P., Thomson, J. A. & Varsani, A. (2007). Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. *J Gen Virol* **88**, 3154-3165.

Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225.

Padilla-Rodriguez, M., Rosario, K. & Breitbart, M. (2013). Novel cyclovirus discovered in the Florida woods cockroach Eurycotis floridana (Walker). *Arch Virol* **158**, 1389-1392.

Park, E. J., Kim, K. H., Abell, G. C., Kim, M. S., Roh, S. W. & Bae, J. W. (2011). Metagenomic analysis of the viral communities in fermented foods. *Appl Environ Microbiol* **77**, 1284-1291.

Parsley, L. C., Consuegra, E. J., Thomas, S. J., Bhavsar, J., Land, A. M., Bhuiyan, N. N., Mazher, M. A., Waters, R. J., Wommack, K. E., Harper, W. F., Jr. & Liles, M. R. (2010). Census of the viral metagenome within an activated sludge microbial assemblage. *Appl Environ Microbiol* **76**, 2673-2677.

Pearson, M. & Pone, S. (1988). Viruses of vanilla in the Kingdom of Tonga. *Australasian Plant Pathology* **17**, 59-60.

Phan, T. G., Kapusinszky, B., Wang, C., Rose, R. K., Lipton, H. L. & Delwart, E. L. (2011). The fecal viral flora of wild rodents. *PLoS Pathog* **7**, e1002218.

Phan, T. G., Luchsinger, V., Avendano, L. F., Deng, X. & Delwart, E. (2014). Cyclovirus in nasopharyngeal aspirates of Chilean children with respiratory infections. *J Gen Virol* **95**, 922-927.

Phan, T. G., Mori, D., Deng, X., Rajindrajith, S., Ranawaka, U., Fan Ng, T. F., Bucardo-Rivera, F., Orlandi, P., Ahmed, K. & Delwart, E. (2015). Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**, 98-104.

Pierson, E. D. & Rainey, W. E. (1992). The biology of flying foxes of the genus Pteropus: a review. In *Pacific island flying foxes: proceedings of an international conservation conference*, pp. 1-17: DTIC Document.

Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *J Gen Virol* **82**, 655-665.

Plowright, R. K., Eby, P., Hudson, P. J., Smith, I. L., Westcott, D., Bryden, W. L., Middleton, D., Reid, P. A., McFarlane, R. A., Martin, G., Tabor, G. M., Skerratt, L. F., Anderson, D. L., Crameri, G., Quammen, D., Jordan, D., Freeman, P., Wang, L. F., Epstein, J. H., Marsh, G. A., Kung, N. Y. & McCallum, H. (2015). Ecological dynamics of emerging bat virus spillover. *Proc Biol Sci* **282**, 20142124.

Quaiser, A., Dufresne, A., Ballaud, F., Roux, S., Zivanovic, Y., Colombet, J., Sime-Ngando, T. & Francez, A. J. (2015). Diversity and comparative genomics of Microviridae in Sphagnum- dominated peatlands. *Front Microbiol* **6**, 375.

Reavy, B., Swanson, M. M., Cock, P. J., Dawson, L., Freitag, T. E., Singh, B. K., Torrance, L., Mushegian, A. R. & Taliansky, M. (2015). Distinct circular single-

stranded DNA viruses exist in different soil types. *Appl Environ Microbiol* **81**, 3934-3945.

Reuter, G., Boros, A., Delwart, E. & Pankovics, P. (2014). Novel circular single-stranded DNA virus from turkey faeces. *Arch Virol* **159**, 2161-2164.

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

Rinke, D. (1991). Birds of'Ata and Late, and additional notes on the avifauna of Niuafo'ou, Kingdom of Tonga. *Notornis* **38**, 131-151.

Rohwer, F., Prangishvili, D. & Lindell, D. (2009). Roles of viruses in the environment. *Environ Microbiol* **11**, 2771-2774.

Rojas, M. R., Hagen, C., Lucas, W. J. & Gilbertson, R. L. (2005). Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. *Annu Rev Phytopathol* **43**, 361-394.

Rosario, K. & Breitbart, M. (2011). Exploring the viral world through metagenomics. *Curr Opin Virol* **1**, 289-297.

Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M. & Varsani, A. (2012a). Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol* **93**, 2668-2681.

Rosario, K., Duffy, S. & Breitbart, M. (2012b). A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* **157**, 1851-1871.

Rosario, K., Marinov, M., Stainton, D., Kraberger, S., Wiltshire, E. J., Collings, D. A., Walters, M., Martin, D. P., Breitbart, M. & Varsani, A. (2011). Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). *J Gen Virol* **92**, 1302-1308.

Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y. & Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol* **11**, 2806-2820.

Rosario, K., Padilla-Rodriguez, M., Kraberger, S., Stainton, D., Martin, D. P., Breitbart, M. & Varsani, A. (2013). Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Epiprocta) from Puerto Rico. *Virus Res* **171**, 231-237.

Rosario, K., Schenck, R. O., Harbeitner, R. C., Lawler, S. N. & Breitbart, M. (2015). Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Front Microbiol* **6**, 696.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E.,

Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T. & Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348-352.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., Colombet, J., Sime-Ngando, T. & Debroas, D. (2012). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One* **7**, e33641.

Sachsenroder, J., Braun, A., Machnowska, P., Ng, T. F., Deng, X., Guenther, S., Bernstein, S., Ulrich, R. G., Delwart, E. & Johne, R. (2014). Metagenomic identification of novel enteric viruses in urban wild rats and genome characterization of a group A rotavirus. *J Gen Virol* **95**, 2734-2747.

Sachsenroder, J., Twardziok, S., Hammerl, J. A., Janczyk, P., Wrede, P., Hertwig, S. & Johne, R. (2012). Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. *PLoS One* **7**, e34631.

Sano, Y., Wada, M., Hashimoto, Y., Matsumoto, T. & Kojima, M. (1998). Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. *J Gen Virol* **79 (Pt 12)**, 3111-3118.

Sasaki, M., Orba, Y., Ueno, K., Ishii, A., Moonga, L., Hang'ombe, B. M., Mweene, A. S., Ito, K. & Sawa, H. (2015). Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses. *J Gen Virol* **96**, 440-452.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G. & Stanley, J. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology* **293**, 63-74.

Savory, F. R. & Ramakrishnan, U. (2014). Asymmetric patterns of reassortment and concerted evolution in Cardamom bushy dwarf virus. *Infect Genet Evol* **24**, 15-24.

Shackelton, L. A., Parrish, C. R., Truyen, U. & Holmes, E. C. (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci U S A* **102**, 379-384.

Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A. & Delwart, E. (2011). The fecal virome of pigs on a high-density farm. *J Virol* **85**, 11697-11708.

Sharman, M., Thomas, J. E., Skabo, S. & Holton, T. A. (2008). Abaca bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Arch Virol* **153**, 135-147.

Shendure, J. & Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145.

Sikorski, A., Massaro, M., Kraberger, S., Young, L. M., Smalley, D., Martin, D. P. & Varsani, A. (2013). Novel myco-like DNA viruses discovered in the faecal matter of various animals. *Virus Res* **177**, 209-216.

Smith, I. & Wang, L. F. (2013). Bats and their virome: an important source of emerging viruses capable of infecting humans. *Curr Opin Virol* **3**, 84-91.

Smits, S. L., Zijlstra, E. E., van Hellemond, J. J., Schapendonk, C. M., Bodewes, R., Schurch, A. C., Haagmans, B. L. & Osterhaus, A. D. (2013). Novel cyclovirus in human cerebrospinal fluid, Malawi, 2010-2011. *Emerg Infect Dis* **19**, 1511.

Spennemann, D. H. & Gary, J. W. G. (2002). An historical attempt to control flying foxes in Samoa with an introduced disease. *Australian Mammalogy* **23**, 177-179.

Stainton, D., Halafihi, M., Collings, D. A. & Varsani, A. (2015a). Genome Sequence of Banana Streak MY Virus from the Pacific Ocean Island of Tonga. *Genome Announc* **3**, e00543-00515.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* **93**, 1103-1119.

Stainton, D., Martin, D. P., Collings, D. A., Thomas, J. E. & Varsani, A. (2016). Identification of defective molecules associated with isolates of Banana bunchy top virus. *Archives of Virology*, DOI: 10.1007/s00705-00015-02736-00700.

Stainton, D., Martin, D. P., Muhire, B. M., Lolohea, S., Lepoint, P., Blomme, G., Crew, K. S., Sharman, M., Kraberger, S. & Dayaram, A. (2015b). The global distribution of Banana bunchy top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events. *Virus Evolution* **1**, vev009.

Steinel, A., Parrish, C. R., Bloom, M. E. & Truyen, U. (2001). Parvovirus infections in wild carnivores. *J Wildl Dis* **37**, 594-607.

Stenlund, A. (2003). Initiation of DNA replication: lessons from viral initiator proteins. *Nat Rev Mol Cell Biol* **4**, 777-785.

Stenzel, T., Piasecki, T., Chrzastek, K., Julian, L., Muhire, B. M., Golden, M., Martin, D. P. & Varsani, A. (2014). Pigeon circoviruses display patterns of recombination, genomic secondary structure and selection similar to those of beak and feather disease viruses. *J Gen Virol* **95**, 1338-1351.

Struebig, M. J., Harrison, M. E., Cheyne, S. M. & Limin, S. H. (2007). Intensive hunting of large flying foxes Pteropus vampyrus natunae in Central Kalimantan, Indonesian Borneo. *Oryx* **41**, 390-393.

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801-812.

Swanson, M. M., Fraser, G., Daniell, T. J., Torrance, L., Gregory, P. J. & Taliansky, M. (2009). Viruses in soils: morphological diversity and abundance in the rhizosphere. *Annals of Applied Biology* **155**, 51-60.

Symonds, E. M., Griffin, D. W. & Breitbart, M. (2009). Eukaryotic viruses in wastewater samples from the United States. *Appl Environ Microbiol* **75**, 1402-1409.

Tamaki, H., Zhang, R., Angly, F. E., Nakamura, S., Hong, P. Y., Yasunaga, T., Kamagata, Y. & Liu, W. T. (2012). Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. *Environ Microbiol* **14**, 441-452.

Tattersall, P. (2006). The evolution of parvovirus taxonomy. *Parvoviruses* **1**, 5-14.

Timchenko, T., de Kouchkovsky, F., Katul, L., David, C., Vetten, H. J. & Gronenborn, B. (1999). A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *J Virol* **73**, 10173-10182.

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

Tucker, K. P., Parsons, R., Symonds, E. M. & Breitbart, M. (2011). Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J* **5**, 822-830.

Ushijima, H., Dairaku, M., Honma, H., Sato, K., Sugiyama, K., Kitamura, T. & Inaba, N. (1990). Incidence of antibodies against human immunodeficiency virus, human T-cell lymphotropic virus type 1, hepatitis B virus, hemorrhagic fever with renal syndrome virus and Chlamydia in Tonga and western Samoa. *Jpn J Med Sci Biol* **43**, 101-109.

van den Brand, J. M., van Leeuwen, M., Schapendonk, C. M., Simon, J. H., Haagmans, B. L., Osterhaus, A. D. & Smits, S. L. (2012). Metagenomic analysis of the viral flora of pine marten and European badger feces. *J Virol* **86**, 2360-2365.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y. & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet* **30**, 418-426.

van Doorn, H. R., Nghia, H. D. T., Chau, T. T. H., de Vries, M., Canuti, M., Deijs, M., Jebbink, M. F., Baker, S., Bryant, J. E. & Tham, N. T. (2013). Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *MBio* **4**, e00231-00213.

Varsani, A., Martin, D. P., Navas-Castillo, J., Moriones, E., Hernandez-Zepeda, C., Idris, A., Murilo Zerbini, F. & Brown, J. K. (2014a). Revisiting the classification of curtoviruses based on genome-wide pairwise identity. *Arch Virol* **159**, 1873-1882.

Varsani, A., Monjane, A. L., Donaldson, L., Oluwafemi, S., Zinga, I., Komba, E. K., Plakoutene, D., Mandakombo, N., Mboukoulida, J., Semballa, S., Briddon, R. W., Markham, P. G., Lett, J. M., Lefeuvre, P., Rybicki, E. P. & Martin, D. P. (2009a). Comparative analysis of Panicum streak virus and Maize streak virus diversity, recombination patterns and phylogeography. *Virol J* **6**, 194.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernandez-Zepeda, C., Idris, A., Brown, J. K., Murilo Zerbini, F. & Martin, D. P. (2014b). Establishment of three new

genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch Virol* **159**, 2193-2203.

Varsani, A., Shepherd, D. N., Dent, K., Monjane, A. L., Rybicki, E. P. & Martin, D. P. (2009b). A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. *Virol J* **6**, 36.

Victoria, J. G., Kapoor, A., Li, L., Blinkova, O., Slikas, B., Wang, C., Naeem, A., Zaidi, S. & Delwart, E. (2009). Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol* **83**, 4642-4651.

Wainwright, R. B., McMahon, B. J., Bender, T. R., Heyward, W. L., Nakanishi, S., Wainwright, K. Y., Foliaki, S., Erickson, S. L. & Fields, H. A. (1986). Prevalence of hepatitis B virus infection in Tonga: identifying high risk groups for immunization with hepatitis B vaccine. *Int J Epidemiol* **15**, 567-571.

Wang, Y., Yang, Q. & Wang, Z. (2014). The evolution of nanopore sequencing. *Front Genet* **5**, 449.

Whon, T. W., Kim, M. S., Roh, S. W., Shin, N. R., Lee, H. W. & Bae, J. W. (2012). Metagenomic characterization of airborne viral DNA diversity in the near-surface atmosphere. *J Virol* **86**, 8221-8231.

Wiles, G. J. & Fujita, M. S. (1992). Food plants and economic importance of flying foxes on Pacific islands. In *Pacific island flying foxes: Proceedings of an international conservation conference US Fish Wildl Serv Biol Rep*, pp. 24-35: DTIC Document.

Wiles, G. J. & Payne, N. H. (1986). The Trade in Fruit Bats Pteropus Spp on Guam and Other Pacific Islands. *Biological Conservation* **38**, 143-161.

Womack, A. M., Bohannan, B. J. & Green, J. L. (2010). Biodiversity and biogeography of the atmosphere. *Philos Trans R Soc Lond B Biol Sci* **365**, 3645-3653.

Wong, S., Lau, S., Woo, P. & Yuen, K. Y. (2007). Bats as a continuing source of emerging infections in humans. *Rev Med Virol* **17**, 67-91.

Woo, P. C., Lau, S. K., Teng, J. L., Tsang, A. K., Joseph, M., Wong, E. Y., Tang, Y., Sivakumar, S., Bai, R., Wernery, R., Wernery, U. & Yuen, K. Y. (2014). Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. *Virology* **471-473**, 117-125.

Wu, Z., Ren, X., Yang, L., Hu, Y., Yang, J., He, G., Zhang, J., Dong, J., Sun, L., Du, J., Liu, L., Xue, Y., Wang, J., Yang, F., Zhang, S. & Jin, Q. (2012). Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *J Virol* **86**, 10999-11012.

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S. & Jin, Q. (2015). Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J*.

Xia, H., Hu, C., Zhang, D., Tang, S., Zhang, Z., Kou, Z., Fan, Z., Bente, D., Zeng, C. & Li, T. (2015). Metagenomic profile of the viral communities in Rhipicephalus spp. ticks from Yunnan, China. *PLoS One* **10**, e0121609.

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T. & Takai, K. (2013). Metagenomic analysis of viral communities in (hado)pelagic sediments. *PLoS One* **8**, e57271.

Young, J. C., Chehoud, C., Bittinger, K., Bailey, A., Diamond, J. M., Cantu, E., Haas, A. R., Abbas, A., Frye, L., Christie, J. D., Bushman, F. D. & Collman, R. G. (2015). Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant* **15**, 200-209.

Yu, M. X., Slater, M. R. & Ackermann, H. W. (2006). Isolation and characterization of Thermus bacteriophages. *Archives of Virology* **151**, 663-679.

Yu, N. T., Zhang, Y. L., Feng, T. C., Wang, J. H., Kulye, M., Yang, W. J., Lin, Z. S., Xiong, Z. & Liu, Z. X. (2012). Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* **44**, 488-494.

Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S. A., Li, G., Peng, Y., Xie, J., Cheng, J., Huang, J. & Yi, X. (2010). A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci U S A* **107**, 8387-8392.

Yu, Z.-C., Chen, X.-L., Shen, Q.-T., Zhao, D.-L., Tang, B.-L., Su, H.-N., Wu, Z.-Y., Qin, Q.-L., Xie, B.-B. & Zhang, X.-Y. (2014). Filamentous phages prevalent in Pseudoalteromonas spp. confer properties advantageous to host survival in Arctic sea ice. *The ISME journal*.

Zawar-Reza, P., Arguello-Astorga, G. R., Kraberger, S., Julian, L., Stainton, D., Broady, P. A. & Varsani, A. (2014). Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol* **26**, 132-138.

Zhang, W., Li, L., Deng, X., Kapusinszky, B., Pesavento, P. A. & Delwart, E. (2014). Faecal virome of cats in an animal shelter. *J Gen Virol* **95**, 2553-2564.

Zhong, X., Guidoni, B., Jacas, L. & Jacquet, S. (2015). Structure and diversity of ssDNA Microviridae viruses in two peri-alpine lakes (Annecy and Bourget, France). *Res Microbiol* **166**, 644-654.

Zsak, L., Day, J. M., Oakley, B. B. & Seal, B. S. (2011). The complete genome sequence and genetic analysis of wca82 a novel uncultured microphage from the turkey gastrointestinal system. *Virol J* **8**, 331.

Chapter 2

Cycloviruses, gemycircularviruses and other novel replication-associated protein encoding circular single-stranded viruses in Pacific flying fox (*Pteropus tonganus*) faeces

Contents

2.1 Abstract

Viral metagenomic studies have demonstrated that animal faeces can be a good sampling source for exploring viral diversity associated with the host and its environment. As part of a continuing effort to identify novel circular replication-associated protein encoding single-stranded (CRESS) DNA viruses circulating in the Tongan archipelago, coupled with the fact that bats are a reservoir species of a large number of viruses, I used a metagenomic approach to investigate the CRESS DNA virus diversity in Pacific flying fox (*Pteropus tonganus*) faeces. Faecal matter from four roosting sites located in Ha'avakatolo, Kolovai, Ha'ateiho (ʻAtele) and Lapaha (Takuilau) on Tongatapu Island was collected in April 2014 and January 2015. From these samples I identified five novel cycloviruses representing three putative species, 25 gemycircularviruses representing at least 14 putative species, 17 other CRESS DNA viruses (15 putative species), two circular DNA molecules and a putative novel multi-component virus for which I have identified three cognate molecules. This study demonstrates that there exists a large diversity of CRESS DNA viruses in Pacific flying fox faeces.

2.2 Introduction

Viral metagenomic studies using next generation sequencing have shown that small circular single-stranded DNA (ssDNA) viruses are ubiquitous in nature. A large number of viruses have been identified that share some similarity to proteins encoded by eukaryote-infecting ssDNA viruses in the *Circoviridae*, *Geminiviridae* and *Nanoviridae* families. Members of the genus *Circovirus* in the *Circoviridae* family have small genomes of ~2kb and are known to infect various birds, mammals and fish (King *et al.*, 2011; Lorincz *et al.*, 2011). The majority of circovirus infections do not appear to cause obvious disease symptoms, however, *Porcine circovirus* -2 and *Beak and feather disease virus* (BFDV) cause post weaning multisystemic wasting syndrome and psittacine beak and feather disease, respectively (Morozov *et al.*, 1998; Ritchie *et al.*, 1989). Members of the *Geminiviridae* and *Nanoviridae* families infect plants causing major crop losses throughout the world and are vectored by aphids, leafhoppers, plant hoppers, tree hoppers and whiteflies (King *et al.*, 2011). Most members of the *Geminiviridae* family have ~2.5 - 3kb monopartite genomes, however, some members of the *Begomovirus* genus have bipartite genomes made up of two components (each ~2.5kb) (King *et al.*, 2011; Varsani *et al.*, 2014b). Monopartite begomoviruses are often associated with satellite DNA molecules known as alphasatellites and betasatellites, which can affect pathogenicity and symptomology in the host (Zhou, 2013). Members of the *Nanoviridae* family have multi-component (6-8) genomes, each component is ~1kb and encodes a single protein and each component is packaged into an individual virion (King *et al.*, 2011).

Circoviruses, nanoviruses, geminiviruses and their associated alphasatellites molecules, all encode a replication-associated protein (Rep) which is essential for initiating rolling circle replication (RCR). Conserved motifs within the Rep are important domains for CRESS viruses and molecules to replicate through RCR (Rosario *et al.*, 2012c). These conserved motifs are divided into two main categories based on their roles, the RCR motifs I, II and III and the superfamily 3 (SF3) helicase motifs known as Walker-A, Walker-B and motif C as reviewed in Rosario *et al.* (2012c). Additionally, a conserved motif known as geminivirus Rep sequence (GRS) which is located downstream of RCR motif II has been identified in geminiviruses (Nash *et al.*, 2011). Over the last decade, a large number of novel eukaryotic circular Rep-encoding ssDNA (CRESS DNA) viruses and molecules have been identified whose Reps contain the conserved RCR and SF3 helicase motifs (similar to those of circoviruses, geminiviruses and nanoviruses). The eukaryotic CRESS DNA viruses have been

identified in various environmental samples, including sea water, deep-sea vents and marine organisms, soil, aquifers, fresh water lakes, Antarctic lakes and ponds, hot springs, wastewater and near surface atmosphere (Breitbart *et al.*, 2015; Dayaram *et al.*, 2015a; Diemer & Stedman, 2012; Dunlap *et al.*, 2013; Fahsbender *et al.*, 2015; Hewson *et al.*, 2013a; Hewson *et al.*, 2013b; Kim *et al.*, 2008; Kraberger *et al.*, 2015a; Labonte & Suttle, 2013; Ng *et al.*, 2012; Ng *et al.*, 2013; Phan *et al.*, 2015; Reavy *et al.*, 2015; Rosario *et al.*, 2009; Rosario *et al.*, 2015; Roux *et al.*, 2013; Roux *et al.*, 2012; Smith *et al.*, 2013; Soffer *et al.*, 2014; Whon *et al.*, 2012; Yoshida *et al.*, 2013; Zawar-Reza *et al.*, 2014). Additionally, CRESS DNA viruses have been identified from various insects and plant material (Basso *et al.*, 2015; Dayaram *et al.*, 2014; Dayaram *et al.*, 2012; Dayaram *et al.*, 2013; Dayaram *et al.*, 2015b; Du *et al.*, 2014; Garigliany *et al.*, 2015; Kraberger *et al.*, 2015b; Male *et al.*, 2015; Ng *et al.*, 2011; Padilla-Rodriguez *et al.*, 2013; Pham *et al.*, 2013; Rosario *et al.*, 2012a; Rosario *et al.*, 2011).

Since faecal samples reflect the presence of viruses associated with a given organism, their diet and/or its surrounding environment, it has proved to be a useful non-invasive approach for surveying CRESS DNA viruses in an ecosystem. Faecal samples from various animals has revealed a large diversity of CRESS DNA viruses (Blinkova *et al.*, 2010; Breitbart *et al.*, 2015; Castrignano *et al.*, 2013; Cheung *et al.*, 2014a; b; Cheung *et al.*, 2013; Cheung *et al.*, 2015; Conceicao-Neto *et al.*, 2015; Delwart & Li, 2012; Ge *et al.*, 2012; Hansen *et al.*, 2015; He *et al.*, 2013; Kim *et al.*, 2012; Kraberger *et al.*, 2015a; Li *et al.*, 2015a; Li *et al.*, 2010a; Li *et al.*, 2010b; Li *et al.*, 2011; Li *et al.*, 2015b; Lima *et al.*, 2015; Ng *et al.*, 2014; Ng *et al.*, 2012; Phan *et al.*, 2015; Reuter *et al.*, 2014; Sachsenroder *et al.*, 2012; Sasaki *et al.*, 2015; Shan *et al.*, 2011; Sikorski *et al.*, 2013a; Sikorski *et al.*, 2013b; van den Brand *et al.*, 2012; Varsani *et al.*, 2014a; Varsani *et al.*, 2015; Woo *et al.*, 2014; Wu *et al.*, 2012; Wu *et al.*, 2015; Zhang *et al.*, 2014).

The majority of these novel eukaryotic CRESS DNA viruses cannot be classified within the existing viral taxonomy frame work of *Circoviridae*, *Geminiviridae* and *Nanoviridae* due to the fact that they are highly diverse, have different genome organisations and most important of all, their hosts are unknown. Nonetheless, groupings within these novel CRESS DNA viruses are beginning to emerge as viral databases are being populated with more sequence data. Of these, the two notable ones are the groups cyclovirus proposed by Li *et al.* (2010b) and gemycircularvirus proposed by Rosario *et al.* (2012a).

Li *et al.* (2010b) proposed a genus cyclovirus within the *Circoviridae* family to accommodate cycloviruses which encode two major ORFs, Rep and capsid protein (CP), in an ambisense organisation. The Reps of cycloviruses are most closely related to those of circoviruses. However, unlike circoviruses, the *cp* of cycloviruses is present in the virion sense and the *rep* is on the complementary sense. Cycloviruses also contain a long intergenic region (LIR) between the start codons and either no IR or a shorter IR than circoviruses between the stop codons of the Rep and the CP (Delwart & Li, 2012; Rosario *et al.*, 2012c). Cycloviruses have mainly been identified to be associated with bats, wild animal faeces, farm animals meat products, insects, human cerebrospinal fluid, respiratory secretion, serum and faeces, rodent intestinal content and equine nasal secretions (Dayaram *et al.*, 2013; Garigliany *et al.*, 2014; Ge *et al.*, 2011; Li *et al.*, 2015a; Li *et al.*, 2010a; Li *et al.*, 2010b; Li *et al.*, 2011; Lima *et al.*, 2015; Padilla-Rodriguez *et al.*, 2013; Phan *et al.*, 2014; Phan *et al.*, 2015; Rosario *et al.*, 2012a; Rosario *et al.*, 2011; Sasaki *et al.*, 2015; Sato *et al.*, 2015; Smits *et al.*, 2013; Tan *et al.*, 2013; Wu *et al.*, 2015; Zhang *et al.*, 2014).

The group gemycircularvirus was proposed by Rosario *et al.* (2012a) and these viruses have ~2.2 kb ambisense genomes encoding a *cp* in the virion sense and a *rep* in the complementary sense. The Reps of gemycircularviruses are most similar to those of geminiviruses and have a GRS domain (Dayaram *et al.*, 2012). Of all the identified gemycircularviruses, only Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1) has been associated with a known host, *Sclerotinia sclerotiorum*, inducing hypovirulence (Yu *et al.*, 2010). Gemycircularviruses have been recovered from various animal faeces, cattle and rat serum, bird buccal and cloacal swab, bat pharyngeal and anal swab, human faeces, blood, cervix and cerebrospinal fluid, treated and raw sewage, insects, river sediments and plant material (Conceicao-Neto *et al.*, 2015; Dayaram *et al.*, 2012; Dayaram *et al.*, 2015b; Du *et al.*, 2014; Hanna *et al.*, 2015; Kraberger *et al.*, 2015a; Kraberger *et al.*, 2015b; Kraberger *et al.*, 2013; Lamberto *et al.*, 2014; Li *et al.*, 2015b; Male *et al.*, 2015; Ng *et al.*, 2014; Ng *et al.*, 2011; Phan *et al.*, 2015; Rosario *et al.*, 2012a; Sikorski *et al.*, 2013b; van den Brand *et al.*, 2012; Wu *et al.*, 2015).

Very little is known about CRESS DNA viruses circulating in the Tongan archipelago and to date only a handful of these have been identified. These include a nanovirus (*Banana bunchy top virus*) (Stainton *et al.*, 2012; Stainton *et al.*, 2016; Stainton *et al.*, 2015), a cyclovirus (Dragonfly cyclovirus - 1), two gemycircularviruses (Dragonfly associated circular virus – 3

and Poaceae associated gemycircularvirus - 1) (Male *et al.*, 2015; Rosario *et al.*, 2012b) and two unclassified CRESS DNA viruses (Dragonfly circularisvirus and Dragonfly orbiculatusvirus) (Rosario *et al.*, 2012a; Rosario *et al.*, 2011). Hence I decided to explore the diversity of CRESS associated with Pacific flying foxes (*Pteropus tonganus*) faecal matter as part of this study. Pacific flying foxes are fruit-eating bats and found throughout the Pacific. It is the only bat species found in the Tongan archipelago (Miller & Wilson, 1997). Pacific flying foxes roost in trees and can live up to 30 years. They feed on fruits, leaves and nectar and play a crucial role in pollination and seed dispersal (Banack, 1998; Nelson *et al.*, 2005; Pierson & Rainey, 1992).

Bats are recognised as natural hosts and perhaps reservoirs of a large diversity of both RNA and DNA viruses (~200 viruses of 27 families) (Calisher *et al.*, 2006), some of these viruses are responsible for emerging infectious and zoonotic viruses (Brook & Dobson, 2015; Calisher *et al.*, 2006; Han *et al.*, 2015; Moratelli & Calisher, 2015; Omatsu *et al.*, 2007; Plowright *et al.*, 2015; Smith & Wang, 2013; Wong *et al.*, 2007). Thus it is not surprising that a significant number of CRESS DNA viruses including circoviruses (n=8), cyloviruses (n=18), gemycircularviruses (n=3), and unclassified viruses (n=22) have been previously recovered from bats (Table 2.1).

In this study, I report the identification of five cycloviruses, 25 gemycircularviruses, 17 unclassified CRESS DNA viruses, two circular DNA molecules and a putative novel multi-component virus from the faeces of Pacific flying fox roosting in Tongatapu, the main island of the Tongan archipelago.

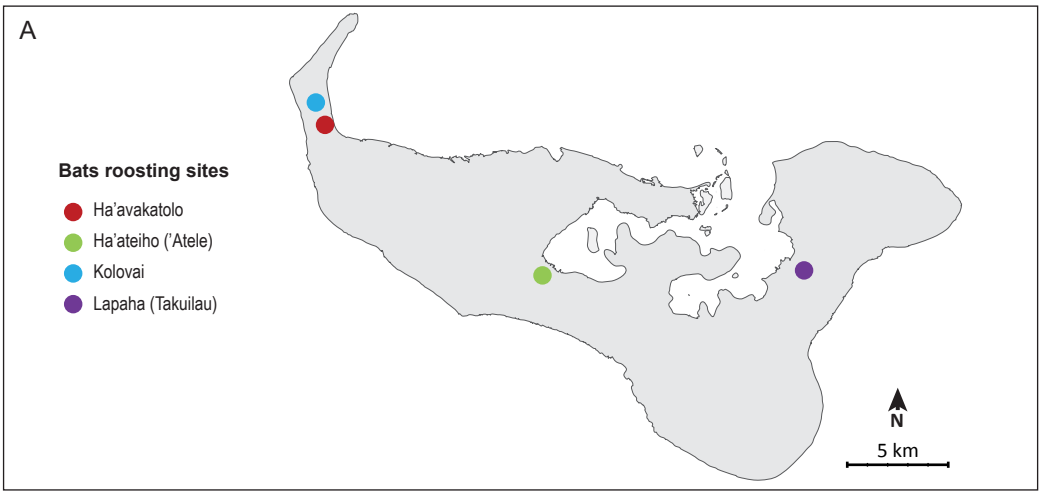
Table 2.1: A summary of all CRESS DNA viruses previously identified associated with bats.

CRESS DNA grouping	Accession	Description	Country	Bat species	Bat diet	Isolation source	Reference
Cyclovirus	JF938079	YN-BtCV-2	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JF938080	YN-BtCV-3	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JF938081	YN-BtCV-4	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JF938082	YN-BtCV-5	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JN377566	Cyclovirus ZS	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	HM228874	GF-4c	USA	*Antrozous pallidus*	Insects	Faeces	Li et al., 2010a
	HQ738637	BaCyV-1	USA	*Tadarida brasiliensis*	Insects	Muscle	Li et al., 2011
	KJ641710	BtMbly-CyV/GS2013	China	*Myotis blythii*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641712	BtRp-CyV-3/GD2012	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641714	BtRp-CyV-14/GD2012	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641715	BtRp-CyV-52/GD2012	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641717	BtMspp.-CyV/GD2012	China	*Myotis* spp.	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641720	BtTp-CyV-2/GX2012	China	*Tylonycteris pachypus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641728	BtPa-CV-2/NX2013	China	*Plecotus auritus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641734	BtVS-CyV/SC2013	China	*Vespertilio superans*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641740	BtRf-CyV-24/YN2010	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KM382269	Bat cyclovirus POA/2012/II	Brazil	*Molossus molossus, Tadarida brasiliensis*	Insects	Faeces	Lima et al.,2015
	KM382270	Bat cyclovirus POA/2012/VI	Brazil	*Molossus molossus, Tadarida brasiliensis*	Insects	Faeces	Lima et al.,2015
Circovirus	JQ814849	RfCV-1	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al., 2012
	JX863737	BtCV XOR1	Myanmar	*Rhinolophus ferrumequinum*	Insects	Stomach contents	He et al., 2013
	KC339249	BtCV XOR7	Myanmar	*Rhinolophus ferrumequinum*	Insects	Stomach contents	He et al., 2013
	KJ641711	BtMr-CV/GD2012	China	*Myotis ricketti*	Fish & water beetles	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641716	BtPspp.-CV/GD2012	China	*Pipistrellus sp.*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641723	BtRs-CV/HuB2013	China	*Rhinolophus sinicus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641724	BtRa-CV/JS2013	China	*Rhinolophus affinis*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641727	BtPa-CV-1/NX2013	China	*Plecotus auritus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
Gemycircularvirus	KJ641719	BtMf-CV-23/GD2012	China	*Miniopterus fuliginosus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641726	BtRf-CV-8/NM2013	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641737	BtRh-CV-6/Tibet2013	China	*Rhinolophus hipposideros*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
Unclassified	JF938078	YN-BtCV-1	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JN377562	Bat circovirus ZS - 00036	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JN377580	Bat circovirus ZS - 00813	China	*Myotis* spp.	Fruit and insects	Faeces	Ge et al., 2011
	JN857329	BTCV-SC703	China	*Insectivorous bat*	Insects	Faeces	Ge et al., 2012
	KJ641713	BtRp-CV-6/GD2012	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641718	BtMf-CV-1/GD2012	China	*Miniopterus fuliginosus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641721	BtTp-CV-3/GX2012	China	*Tylonycteris pachypus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641722	BtMf-CV/HeN2013	China	*Miniopterus fuliginosus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641725	BtRf-CV-1/NM2013	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641729	BtPa-CV-3/NX2013	China	*Plecotus auritus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641730	BtMl-CV/QH2013	China	*Murina leucogaster*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641731	BtRp-CV/SD2013	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641732	BtRf-CV/SX2013	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641733	BtMf-CV/SAX2011	China	*Miniopterus fuliginosus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641735	BtRh-CV-1/Tibet2013	China	*Rhinolophus hipposideros*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641736	BtRh-CV-5/Tibet2013	China	*Rhinolophus hipposideros*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641738	BtRh-CV-7/Tibet2013	China	*Rhinolophus hipposideros*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641739	BtRf-CV-1/YN2010	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641741	BtRf-CV-61/YN2010	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KJ641742	BtRf-CV-62/YN2010	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
	KM382271	Bat circovirus POA/2012/I	Brazil	*Molossus molossus, Tadarida brasiliensis*	Insects	Faeces	Lima et al.,2015
	KM382272	Bat circovirus POA/2012/V	Brazil	*Molossus molossus, Tadarida brasiliensis*	Insects	Faeces	Lima et al.,2015

2.3 Materials and methods

2.3.1 Sample collection and viral DNA isolation

Fresh faecal samples of *P. tonganus* were collected from four bat roosting sites (Ha'ateiho ('Atele), Lapaha (Takuilau), Ha'avakatolo and Kolovai) located on the main Island (Tongatapu) of the pacific archipelago of Tonga in April 2014 and January 2015 (Figure 2.1). Samples were stored at -20°C prior to processing. Samples (~5-10 g) were subsequently thawed, resuspended in 45 ml of SM Buffer (50 mM Tris·HCl, 10 mM $MgSO_4$, 0.1 M NaCl, pH 7.5) by vigorous shaking. The homogenates from each sample were filtered sequentially through 0.45 μm and 0.2 μm syringe filters. The filtrate was precipitated overnight with 15% (w/v) PEG 8000 at 4°C and centrifuged at 14,800g for 10 min. The pellet was resuspended in 1 ml of SM buffer and 200 μl of this was used for viral nucleic acid extraction using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA). Circular viral DNA was enriched by rolling circle amplification using TempliPhi (GE Healthcare, USA).



B

GenBank accession	Viral / DNA molecule sequence name	Acronym	Composite (4 sites)	Ha'avakatolo	Ha'ateiho ('Atele)	Kolovai	Lapaha (Takuilau)
			2014		2015		
KT732785	Pacific flying fox associated cyclovirus-1	PfffaCyV-1		1			
KT732786	Pacific flying fox associated cyclovirus-2	PfffaCyV-2		1			
KT732787 - KT732789	Pacific flying fox associated cyclovirus-3	PfffaCyV-3		1	1	1	
KT732790 - KT732791	Pacific flying fox faeces associated gemycircularvirus-1	PfffaGmV-1		1	1		
KT732792 - KT732793	Pacific flying fox faeces associated gemycircularvirus-2	PfffaGmV-2	1		1		
KT732794	Pacific flying fox faeces associated gemycircularvirus-3	PfffaGmV-3		1			
KT732795 - KT732796	Pacific flying fox faeces associated gemycircularvirus-4	PfffaGmV-4	1	1			
KT732797	Pacific flying fox faeces associated gemycircularvirus-5	PfffaGmV-5			1		
KT732798 - KT732799	Pacific flying fox faeces associated gemycircularvirus-6	PfffaGmV-6		1	1		
KT732800	Pacific flying fox faeces associated gemycircularvirus-7	PfffaGmV-7		1			
KT732801 - KT732802	Pacific flying fox faeces associated gemycircularvirus-8	PfffaGmV-8	1	1			
KT732803	Pacific flying fox faeces associated gemycircularvirus-9	PfffaGmV-9	1				
KT732804 - KT732805	Pacific flying fox faeces associated gemycircularvirus-10	PfffaGmV-10	2				
KT732807 - KT732812	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	1	2	2		2
KT732813	Pacific flying fox faeces associated gemycircularvirus-12	PfffaGmV-12			1		
KT732814	Pacific flying fox faeces associated gemycircularvirus-13	PfffaGmV-13		1			
KT732806	Pacific flying fox faeces associated gemycircularvirus-14	PfffaGmV-14	1				
KT732820 - KT732821	Pacific flying fox faeces associated circular DNA virus-1	PfffaCV-1		1	1		
KT732829, KT732831	Pacific flying fox faeces associated circular DNA virus-2	PfffaCV-2		1	1		
KT732818	Pacific flying fox faeces associated circular DNA virus-3	PfffaCV-3	1				
KT732819	Pacific flying fox faeces associated circular DNA virus-4	PfffaCV-4	1				
KT732822	Pacific flying fox faeces associated circular DNA virus-5	PfffaCV-5		1			
KT732823	Pacific flying fox faeces associated circular DNA virus-6	PfffaCV-6		1			
KT732824	Pacific flying fox faeces associated circular DNA virus-7	PfffaCV-7		1			
KT732825	Pacific flying fox faeces associated circular DNA virus-8	PfffaCV-8		1			
KT732784	Pacific flying fox faeces associated circular DNA virus-9	PfffaCV-9			1		
KT732827	Pacific flying fox faeces associated circular DNA virus-10	PfffaCV-10			1		
KT732828	Pacific flying fox faeces associated circular DNA virus-11	PfffaCV-11		1			
KT732830	Pacific flying fox faeces associated circular DNA virus-12	PfffaCV-12		1			
KT732832	Pacific flying fox faeces associated circular DNA virus-13	PfffaCV-13			1		
KT732833	Pacific flying fox faeces associated circular DNA virus-14	PfffaCV-14	1				
KT732834	Pacific flying fox faeces associated circular DNA virus-15	PfffaCV-15	1				
KT732815 - KT732817	Pacific flying fox associated multicomponent virus-1	PffaMCV-1				1	
KT732783	Pacific flying fox faeces associated circular DNA molecule-1	PffaCM-1		1			
KT732826	Pacific flying fox faeces associated circular DNA molecule-2	PffaCM-2			1		

Figure 2.1: A. Pacific flying fox faeces sampling sites on Tongatapu Island, Tonga. **B.** Summary of viruses recovered from various sites and sampling periods. Numbers in filled coloured circles indicate the number of isolates of the CRESS DNA virus / molecules from each site.

2.3.2 Sequencing and recovery of complete viral genomes

The enriched DNA was sequenced at Beijing Genomics Institute (Hong Kong) on an Illumina HiSeq 2000 sequencer (Illumina, USA). The paired-end reads were *de novo* assembled using ABySS 1.5.2. with a k-mer setting of 64 (Simpson *et al.*, 2009) and the resulting >500 nt contigs were analysed using BLASTx (Altschul *et al.*, 1990) against a viral protein database.

For contigs with hits to proteins encoded by CRESS DNA viruses, abutting primers (Table 2.2) were designed to recover the complete circular DNA molecules by polymerase chain reaction (PCR) using KAPA Hotstart HiFi DNA polymerase (Kapa Biosystems, USA). The resulting PCR amplicons were gel purified, ligated into pJET1.2 plasmid vector (Thermo Fisher Scientific, USA), and the recombinant plasmids were Sanger sequenced using primer walking at Macrogen Inc. (Korea). The Sanger sequence reads were assembled using DNA Baser V4 (Heracle Biosoft S.R.L. Romania).

2.3.3 Sequence analyses

Putative CP and Rep ORFs were identified using ORF Finder coupled with BLASTx (Altschul *et al.*, 1990) analysis. Pairwise similarity comparisons of nucleotide and protein sequences were carried out using SDT v1.2 (Muhire *et al.*, 2014). Sequences of cycloviruses and gemycircularviruses were downloaded from GenBank on the 1st of October 2015. The Reps and CPs encoded by these together with those from viral genomes identified in this study were aligned using PROMALS3D (Pie *et al.*, 2008). These alignments were used to infer maximum-likelihood phylogenetic trees using PHYML (Guindon *et al.*, 2010) and best fit substitution models determined using ProtTest (Abascal *et al.*, 2005) (cycloviruses – Rep: LG+I+G, CP: Blosum62+G+F; gemycircularviruses – Rep: LG+G+I, CP: LG+G+I) with approximate likelihood branch support (aLRT). Branches with less than 80% support were collapsed.

Table 2.2: Sequences of forward and reverse back-to-back primers used to recover complete genomes of CRESS DNA viruses and circular DNA molecules in this study.

Sequence ID	Forward	Reverse
PfffaCyV-1	5'-GGTACTGATGACCAAAATCGCGACTAC-3'	5'-CTTTGCTCCTTCCAGGTGAGCCCTAC-3'
PfffaCyV-2	5'-CTCGAGAAATGGTGGGAGGAAGTGG-3'	5'-CCGTATTAAACCGTATAGCTCGTCCG-3'
PfffaCyV-3	5'-GAGAGCGAGACTTTAAGACCGAAGTG-3'	5'-CACTTCCATGAGCGATTCGTATGTACTG-3'
PfffaGmV-1	5'-CTGGCGTTGGAGATGTGTATGTCTATG-3'	5'-GCTTTCCATCCGTAGAAGTGTAGAATCC-3'
PfffaGmV-2	5'-GTCATATCTACTGCATTGGGCTAGTCTCG-3'	5'-GTCCTAGGGAACGAGCCCACAAG-3'
PfffaGmV-3	5'-GAAGTACAAGCATTGGCATGCCATGG-3'	5'-CGAATGGAGCCATCGTCATTCCC-3'
PfffaGmV-4	5'-CGGAGGCTAAGTATGCTGTCTTCGAC-3'	5'-GCGCATCACGCAGGAGAAGCTTG-3'
PfffaGmV-5	5'-CCTGACTTTGATTGGATGGAGGGC-3'	5'-CCGTGCGTTATTGGGGTAAGAGTCC-3'
PfffaGmV-6	5'-GATTACGTGGCCAAGCATGCAGGC-3'	5'-CCAGCCTCGTTGAGGTTTAGCTCG-3'
PfffaGmV-7	5'-CACGACATCGATTGGGACTGGATG-3'	5'-GTGTAGTGGGTCTTCCCTTGGATCC-3'
PfffaGmV-8	5'-GCACTCCAGCGCAAGCTTATGACTAC-3'	5'-GGCCAATAGGCTTGATGTTTGGGTGC-3'
PfffaGmV-9	5'-GTCTCCTCACTACTTTTCGTGCCAG-3'	5'-CGGTGACCTCAACGAAATCGCAAG-3'
PfffaGmV-10	5'-GTACTCTACAAGGAGCCTGCACTCTAC-3'	5'-CTTCAGCTGAAATTGAGCCTGACATCCC-3'
PfffaGmV-11*	5'-CCAAGTACAAGCGGAAGTACAAGGCC-3'	5'-CCTTGTACCGGTTGGCTTGCTTGG-3'
PfffaGmV-11**	5'-GACTTGTCACCGAGCGAAGTTACATGG-3'	5'-GTAGCGCGGATCCGAATTGCTCAAC-3'
PfffaGmV-12	5'-CTGAATGCATCATCGGAAGAGAGGATC-3'	5'-CTGCAAGTTCAGCAAGATGGTTGACC-3'
PfffaGmV-13	5'-GAGAATGTATCGTGGGAAGAGAGCTAC-3'	5'-CTTGAAGGCCAGATAGGAGGTCCAC-3'
PfffaGmV-14	5'-CCAAGTACAAGCGGAAGTACAAGGCC-3'	5'-CCTTGTACCGGTTGGCTTGCTTGG-3'
PfffaCV-1	5'-CGAAAGCTGACCTGCAGGCTAATCG-3'	5'-CACCCTCAAAGTGAGCTGCTTCAAAGC-3'
PfffaCV-2	5'-CGAGCCGTTAACACCACTCGAGTTTC-3'	5'-CTCGAATTCAATGGCGACGTGAACATG-3'
PfffaCV-3	5'-CACACCACACCTACAGGGGTTCG-3'	5'-CCGCATTCTCCAACCTCCTTCCC-3'
PfffaCV-4	5'-CTGAAACGAAAGGAGGCAACTCTATCATCC-3'	5'-CTTCAAACGACCAGCGATCGGCC-3'
PfffaCV-5	5'-CGTATTGTTGCAAAGACGATGGACCGAG-3'	5'-CAACATTGCGTACCCATAACTTCCGC-3'
PfffaCV-6	5'-GATATGTGGACCTCCTGGTATAGG-3'	5'-CAAACGCCACAGGGCTCACTTAAAC-3'
PfffaCV-7	5'-GATGATGACACCGTGGCTGGAGCTG-3'	5'-GTACAGTTCTCGTCCACGTTTTCCAC-3'
PfffaCV-8	5'-CTGGTGGGACGGATACGAGTATCAAC-3'	5'-GTGCCGCGTTGCTTATAGAAGATGG-3'
PfffaCV-9	5'-GACTGCTGGAAATGCCCATTTTCCAAGG-3'	5'-CTGTACCAGTTCCTCCACTCATAGTTG-3'
PfffaCV-10	5'-GACACATTCAGGAGACGTTGGAACCTG-3'	5'-CGAGTATTGTTGCTTGGCCGATACGC-3'
PfffaCV-11	5'-CTTTCGCCGACTATGTTTCCGAAGTTCC-3'	5'-GTGTACTGTCATCGGGGTTGTTAAGTG-3'
PfffaCV-12	5'-CTTATTGTGCGCTGCGATATTCGGAATG-3'	5'-GCACACAACAGCGTGAATGTGAAG-3'
PfffaCV-13	5'-GAAGACTCTCTGAGCCAGTCATCATCG-3'	5'-CGAGCTGCTTTCCACAGTAGTCAAGG-3'
PfffaCV-14	5'-CGTCCGGATACATTGATTTCAGCCAATCG-3'	5'-GATAACAATATCTGGCCGTGTCGTTTCTGTC-3'
PfffaCV-15	5'-CTCTCGCAACAAGAGGATGGTATGGAAAC-3'	5'-GTGCTCTAGGACCTCCACCCATTC-3'
PfffaCM-1	5'-CGAGAAATGCAGCGAGACGTAC-3'	5'-GCGATGTCTCTCATGGGTGCCTTC-3'
PfffaCM-2	5'-CGTATTGTTGCAAAGACGATGGACCGAG-3'	5'-CAACATTGCGTACCCATAACTTCCGC-3'
PfffaMCV-1 (DNA-R)	5'-GAAACATCTCCAAGGCTACTGCTACTTC-3'	5'-GTACCTTCTTCACCGATTTCTCTTGCCAC-3'
PfffaMCV-1 (DNA-S)	5'-GATACTATTCAGGGTACAATCGCTGCC-3'	5'-GGAATACTTTACCTCAGGGCGTGG-3'
PfffaMCV-1 (DNA-U1)	5'-CCTCCACATTTGACCGATCTTTACGATG-33	5'-GGGAACTTGATTGGAAGGGCACAG-3'

* Primer pair used to recover isolates Tbat_A_103746, Tbat_I_103746, Tbat_H_103746

** Primer pair used to recover isolates Tbat_A_103909, Tbat_I_103909, Tbat_H_103909

2.4 Results and Discussion

2.4.1 Recovery and characterisation of novel CRESS DNA viruses and DNA molecules in Pacific flying fox faeces

I sampled viral genomes from Pacific flying fox faeces from four roosting sites situated on Tongatapu Island of the Tongan archipelago (Figure 2.1). Through a next-generation sequencing informed approach, I identified 601 of the 1119 *de novo* assembled contigs which were >500nts that had viral sequence-like BLAST hits. For the purpose of this study, I concentrated on contigs that had hits to eukaryotic CRESS DNA viral sequences for complete characterisation. Using abutting primers (Table 2.2) that were designed to recover individual CRESS DNA molecules based on the viral-like contigs, I amplified, cloned and Sanger sequenced 48 circular molecules that encode a Rep and an additional large open reading frame (ORF) which putatively encodes a CP (Figure 2.2). I also recovered two circular DNA molecules, one that encodes a bacterial Rep-like element and the other encodes CP-like element. These may be 'subgenomic' molecules or part of a multi-component viruses or possibly non-viral mobile genetic elements. Finally, I also recovered a small Rep-encoding molecule (1159 nts), that is similar in size to the multi-component nanovirus Rep encoding molecules. A similarity search of the non-coding region of this molecule in the bat faeces contigs database revealed two additional molecules with high nucleotide identity (>95%; Figure 2.3) for which I designed abutting primers and recovered these molecules (Table 2.2).

Figure 2.2: Genome organisations of the cycloviruses, gemycircularviruses, circular DNA molecules and unclassified CRESS DNA viruses recovered from Pacific flying fox faeces. *rep*: replication associated protein gene; *cp*: capsid protein gene.

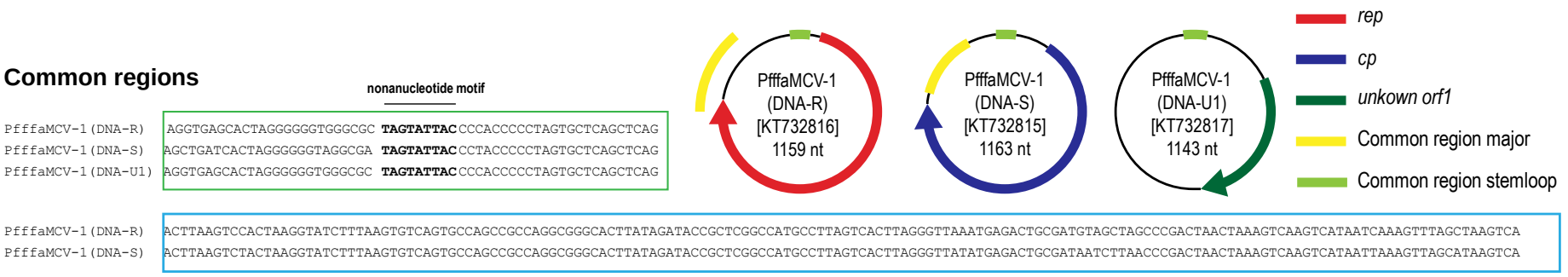


Figure 2.3: Genome organisations of the three components of Pacific flying fox faeces associated multi-component virus-1(PfffaMCV-1) including two common regions.

The CRESS DNA viruses (n=48) I recovered include cycloviruses (n=5; 3 species), gemycircularviruses (n=25; 14 species), unclassified viruses (n=17) and a putative multi-component virus (3 components identified) (Figures 2.1-2.3, Table 2.3). Approximately 40% of the CRESS DNA viruses were recovered from the Ha'avakatolo Pacific flying fox roosting site (Figure 2.1), ~30% Ha'ateiho ('Atele) and two each from Kolovai and Lapaha (Takuilau) (Table 2.3). Interestingly, I would have expected the Kolovai Pacific flying fox roosting colony to have a similar assemblage of viruses as Ha'avakatolo given their close proximity, however, although one viral species (PfffaGmV-11) was present in both Kolovai and Ha'avakatolo, the other viral species identified in Kolovai (PfffaMCV-1) was not found in Ha'avakatolo or in fact in any of the other sites. Instead Ha'avakatolo and Ha'ateiho which are geographically more distant share six of the same virus species. Also of note, in three cases the viruses sampled in 2014 were also identified in 2015 suggesting that these viruses may be persistently associated with Pacific flying fox faecal matter. Furthermore, six viruses were found at two or more sampling sites in 2015 suggesting that these viruses are possibly common in Pacific flying fox roosting colonies. It is unknown whether the Pacific flying foxes in Tonga move between roosting sites or roost at specific colonies through their lifetime.

Table 2.3: Summary of all the viruses and DNA molecules recovered from Pacific flying fox faeces in this study.

GenBank Accession #	Isolate	Sampling site in Tonga	Name	Lengths (nts)	Sampling year	Genetic code
KT732785	Tbat_H_103699	Ha'avakatolo	Pacific flying fox associated cyclovirus-1	1923	2015	Standard
KT732786	Tbat_H_88317	Ha'avakatolo	Pacific flying fox associated cyclovirus-2	1916	2015	Standard
KT732787	Tbat_K_103923	Kolovai	Pacific flying fox associated cyclovirus-3	1838	2015	Standard
KT732788	Tbat_H_103923	Ha'avakatolo	Pacific flying fox associated cyclovirus-3	1838	2015	Standard
KT732789	Tbat_A_103923	Ha'ateiho ('Atele)	Pacific flying fox associated cyclovirus-3	1838	2015	Standard
KT732820	Tbat_A_1180	Ha'ateiho ('Atele)	Pacific flying fox faeces associated circular DNA virus-1	2254	2015	Standard
KT732821	Tbat_H_1180	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-1	2254	2015	Standard
KT732829	Tbat_A_103763	Ha'ateiho ('Atele)	Pacific flying fox faeces associated circular DNA virus-2	2538	2015	Standard
KT732831	Tbat_H_103763	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-2	2538	2015	Standard
KT732818	Tbat_38855	4 sites combined	Pacific flying fox faeces associated circular DNA virus-3	2214	2014	Standard
KT732819	TBAT_29894	4 sites combined	Pacific flying fox faeces associated circular DNA virus-4	1757	2014	Standard
KT732822	Tbat_H_25288	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-5	2694	2015	Standard
KT732823	Tbat_H_77994	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-6	1963	2015	Standard
KT732824	Tbat_H_85975	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-7	2722	2015	Standard
KT732825	Tbat_H_103163	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-8	2004	2015	Standard
KT732784	Tbat_A_77299	Ha'ateiho ('Atele)	Pacific flying fox faeces associated circular DNA virus-9	2707	2015	Ciliate
KT732827	Tbat_A_16877	Ha'ateiho ('Atele)	Pacific flying fox faeces associated circular DNA virus-10	2522	2015	Standard
KT732828	Tbat_H_102636	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-11	1947	2015	Standard
KT732830	Tbat_H_65519	Ha'avakatolo	Pacific flying fox faeces associated circular DNA virus-12	2655	2015	Standard
KT732832	Tbat_A_103819	Ha'ateiho ('Atele)	Pacific flying fox faeces associated circular DNA virus-13	2630	2015	Standard
KT732833	Tbat_5606	4 sites combined	Pacific flying fox faeces associated circular DNA virus-14	2732	2014	Standard
KT732834	Tbat_3598	4 sites combined	Pacific flying fox faeces associated circular DNA virus-15	2571	2014	Standard
KT732790	Tbat_A_103952	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-1	2233	2015	Standard
KT732791	Tbat_H_103952	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-1	2233	2015	Standard
KT732792	Tbat_103791	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-2	2250	2014	Standard
KT732793	Tbat_A_103791	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-2	2250	2015	Standard
KT732794	Tbat_H_103958	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-3	2209	2015	Standard
KT732795	TBAT_21383	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-4	2140	2014	Standard
KT732796	Tbat_H_103639	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-4	2145	2015	Standard
KT732797	Tbat_A_103852	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-5	2172	2015	Standard
KT732798	Tbat_A_103779	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-6	2220	2015	Standard
KT732799	Tbat_H_103779	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-6	2219	2015	Standard
KT732800	Tbat_H_103921	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-7	2217	2015	Standard
KT732801	Tbat_12377	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-8	2205	2014	Standard
KT732802	Tbat_H_12377	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-8	2205	2015	Standard
KT732803	Tbat_103951	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-9	2230	2014	Standard
KT732804	Tbat_45285	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-10	2191	2014	Standard
KT732805	Tbat_47364	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-10	2189	2014	Standard
KT732807	Tbat_A_103746	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-11	2155	2015	Standard
KT732808	Tbat_A_103909	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-11	2159	2015	Standard
KT732809	Tbat_H_103746	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-11	2155	2015	Standard
KT732810	Tbat_H_103909	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-11	2156	2015	Standard
KT732811	Tbat_L_103746	Lapaha (Takuilau)	Pacific flying fox faeces associated gemycircularvirus-11	2156	2015	Standard
KT732812	Tbat_L_103909	Lapaha (Takuilau)	Pacific flying fox faeces associated gemycircularvirus-11	2159	2015	Standard
KT732813	Tbat_A_64418	Ha'ateiho ('Atele)	Pacific flying fox faeces associated gemycircularvirus-12	2310	2015	Standard
KT732814	Tbat_H_103806	Ha'avakatolo	Pacific flying fox faeces associated gemycircularvirus-13	2250	2015	Standard
KT732806	Tbat_31579	4 sites combined	Pacific flying fox faeces associated gemycircularvirus-14	2185	2014	Standard
KT732815	Tbat_K_12099_CP	Kolovai	Pacific flying fox associated multicomponent virus-1	1163	2015	Standard
KT732816	Tbat_K_12099_Rep	Kolovai	Pacific flying fox associated multicomponent virus-1	1159	2015	Standard
KT732817	Tbat_K_12099_unk	Kolovai	Pacific flying fox associated multicomponent virus-1	1143	2015	Standard
KT732783	Tbat_H_67299	Ha'avakatolo	Pacific flying fox faeces associated circular DNA molecule-1	1957	2015	Standard
KT732826	Tbat_A_25288	Ha'ateiho ('Atele)	Pacific flying fox faeces associated circular DNA molecule-2	2255	2015	Standard

2.4.2 Cycloviruses

Five cyclovirus genomes (KT732785 - KT732789) were recovered in this study from Pacific flying fox faeces. Based on the proposed guidelines of 80% full genome pairwise identity species cut-off (see cyclovirus proposal at ICTV), I have classified these five cycloviruses into three putative species, which have been tentatively named Pacific flying fox faeces-associated cyclovirus (PfffaCyV) 1 to 3; PfffaCyV-1 (n=1), PfffaCyV-2 (n=1) and PfffaCyV-3 (n=3) (Figure 2.1B, 2.2; Table 2.3). The long intergenic region (LIR) of PfffaCyV contains the putative origin of replication (*ori*) with the conserved nonanucleotide motif (TAGTATTAC) at the apex of a stem-loop structure. All the PffaCyVs appear to have putative spliced Reps similar to those found in a subset of cycloviruses (GenBank accession #s: AB937980 - AB937987, GQ404857 - GQ404858, HQ738634 - HQ738635, JX185424, JX569794, KC771281, KF031465 - KF031471, KM392284 - KM392289; Table 2.4).

Analysis of the nucleotide pairwise identities of the cyclovirus sequences from this study together with those available in GenBank using SDT v1.2 (Muhire *et al.*, 2014) revealed that there is ~47% diversity within the entire cyclovirus group (n=109). The genome sequences of the three PfffaCyV-3s (KT732787 - KT732789) share >99% identity. The genomes of PfffaCyV-2 (KT732786) and PfffaCyV-1 (KT732785) share 79% and ~60% pairwise identity respectively with those of PfffaCyV-3s. Genome-wide percentage pairwise identities of cycloviruses generated using SDT v1.2 (Muhire *et al.*, 2014) are provided in Supplementary Data 2.1. The Reps and CPs of PfffaCyV-1 and PfffaCyV-2 share 75% and 67% pairwise amino acid identity, respectively. The PfffaCyV-1 and PfffaCyV-2 Reps share ~47% amino acid identity with those of PfffaCyV-3s. The maximum-likelihood phylogenetic trees of the cyclovirus Rep and CP amino acid sequences (Figure 2.4) show that the PfffaCyV-3 Reps and CPs are most closely related to those of Human cyclovirus VS5700009 (KC771281), recovered from blood serum and cerebrospinal fluid of patients with unexplained paraplegia in Malawi (Smits *et al.*, 2013). PfffaCyV-3 Reps and CPs share ~90% and 56% amino acid identity respectively to those of Human cyclovirus VS5700009. PfffaCyV-3s share 78% genome-wide identity with that of Human cyclovirus VS5700009. It is also worth noting that PfffaCyV-3 was recovered from three Pacific flying fox roosting sites in 2015 (Figure 2.1) indicating that this virus is commonly circulating in these colonies in Tongatapu.

The Reps of PfffaCyV-1, -2 and -3 contain all motifs that are conserved in other cycloviruses as reviewed in Rosario *et al.* (2012c) with the exceptions of motif III [**Y**Cx/S**K**] where PfffaCyV-1 and PfffaCyV-2 contain a H residue instead of the conserved K, and a L residue instead of C (Table 2.5).

Table 2.4: Summary of all known cycloviruses as of 1st October 2015.

Genbank Accession	Cycloviruses description	Acronyms	Country	Isolation source	Common name	Sample type	References
AB937980	Cyclovirus ZM32	ZM32	Zambia	*Mastomys natalensis*	African rat	Faeces	Sasaki et al., 2015
AB937981	Cyclovirus ZM01	ZM01	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
AB937982	Cyclovirus ZM36a	ZM36a	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
AB937983	Cyclovirus ZM38	ZM38	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
AB937984	Cyclovirus ZM41	ZM41	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
AB937985	Cyclovirus ZM50a	ZM50a	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
AB937986	Cyclovirus ZM54	ZM54	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
AB937987	Cyclovirus ZM62	ZM62	Zambia	*Crocidura hirta*	Lesser red musk shrew	Faeces	Sasaki et al., 2015
GQ404844	Cyclovirus PK5006	PK5006	Pakistan	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404845	Cyclovirus PK5034	PK5034	Pakistan	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404846	Cyclovirus PK5222	PK5222	Pakistan	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404847	Cyclovirus PK5510	PK5510	Pakistan	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404848	Cyclovirus PK6197	PK6197	Pakistan	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404849	Cyclovirus Chimp11	Chimp11	Central Africa	*Pan troglodytes*	African chimpanzee	Faeces	Li et al., 2010b
GQ404850	Cyclovirus Chimp12	Chimp12	Central Africa	*Pan troglodytes*	African chimpanzee	Faeces	Li et al., 2010b
GQ404854	Cyclovirus NG12	NG12	Nigeria	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404855	Cyclovirus NG14	NG14	Nigeria	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404857	Cyclovirus TN25	TN25	Tunisia	*Homo sapiens*	Human	Faeces	Li et al., 2010b
GQ404858	Cyclovirus TN18	TN18	Tunisia	*Homo sapiens*	Human	Faeces	Li et al., 2010b
HM228874	Bat cyclovirus GF-4c	GF-4c	USA	*Antrozous pallidus*	Bat	Faeces	Li et al., 2010a
HQ638049	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Tholymis tillarga*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638050	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Tholymis tillarga*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638051	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Tholymis tillarga*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638052	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638053	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638054	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638055	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638056	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638057	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Tholymis tillarga*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638058	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638059	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638060	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638061	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638062	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638063	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638064	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011

Genbank Accession	Cycloviruses description	Acronyms	Country	Isolation source	Common name	Sample type	References
HQ638065	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638066	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638067	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638068	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Diplacodes bipunctata*	Dragonfly	Abdomen	Rosario et al., 2011
HQ638069	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Tholymis tillarga*	Dragonfly	Abdomen	Rosario et al., 2011
HQ738634	Cyclovirus PKbeef23	PKbeef23	Pakistan	*Bos taurus*	Cow	Muscle	Li et al., 2011
HQ738635	Cyclovirus PKgoat21	PKgoat21	Pakistan	*Bos taurus*	Cow	Muscle	Li et al., 2011
HQ738636	Cyclovirus PKgoat11	PKgoat11	Pakistan	*Capra aegagrus hircus*	Goat	Muscle	Li et al., 2011
HQ738637	Cyclovirus CyCV-TB	CyCV-TB	USA	*Tadarida brasiliensis*	Bat	Muscle	Li et al., 2011
HQ738643	Cyclovirus NGchicken8	NGchicken8	Nigeria	*Gallus gallus*	Chicken	Muscle	Li et al., 2011
HQ738644	Cyclovirus NGchicken15	NGchicken15	Nigeria	*Gallus gallus*	Chicken	Muscle	Li et al., 2011
JF938079	Bat circovirus ZS/China/2011*	YN-BtCV-2	China	*Myotis* spp.	Bat	Faeces	Ge et al., 2011
JF938080	Bat circovirus ZS/China/2011*	YN-BtCV-3	China	*Myotis* spp.	Bat	Faeces	Ge et al., 2011
JF938081	Bat circovirus ZS/China/2011*	YN-BtCV-4	China	*Myotis* spp.	Bat	Faeces	Ge et al., 2011
JF938082	Bat circovirus ZS/China/2011*	YN-BtCV-5	China	*Myotis* spp.	Bat	Faeces	Ge et al., 2011
JN377566	Bat circovirus ZS/Yunnan-China/2009*	Cyclovirus ZS	China	*Myotis* spp.	Bat	Faeces	Ge et al., 2011
JX185419	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185420	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185421	Dragonfly cyclovirus 1	DfCyV-1	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185422	Dragonfly cyclovirus 2	DfCyV-2	USA	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185423	Dragonfly cyclovirus 2	DfCyV-2	USA	*Anax junius*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185424	Dragonfly cyclovirus 3	DfCyV-3	USA	*Erythemis simplicicollis*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185425	Dragonfly cyclovirus 4	DfCyV-4	Bulgaria	*Somatochlora meridionalis*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185426	Dragonfly cyclovirus 5	DfCyV-5	Puerto Rico	*Erythrodiplax umbrata*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185427	Dragonfly cyclovirus 5	DfCyV-5	Puerto Rico	*Erythrodiplax umbrata*	Dragonfly	Abdomen	Rosario et al., 2012a
JX569794	Florida woods cockroach-associated cyclovirus	FWCasCyV-1	USA	*Eurycotis floridana*	Florida wood cockroach	Abdomen	Padilla-Rodriguez et al., 2013
KC512916	Dragonfly cyclovirus 4	DfCyV-4	USA	*Aeshna multicolor*	Dragonfly	Abdomen	Dayaram et al., 2013
KC512917	Dragonfly cyclovirus 4	DfCyV-4	USA	*Aeshna multicolor*	Dragonfly	Abdomen	Dayaram et al., 2013
KC512918	Dragonfly cyclovirus 6	DfCyV-6	USA	*Aeshna multicolor*	Dragonfly	Abdomen	Dayaram et al., 2013
KC512919	Dragonfly cyclovirus 7	DfCyV-7	New Zealand	*Xanthocnemis zealandica*	Dragonfly	Abdomen	Dayaram et al., 2013
KC512920	Dragonfly cyclovirus 8	DfCyV-8	Australia	*Orthetrum sabina*	Dragonfly	Abdomen	Dayaram et al., 2013
KC771281	Human cyclovirus VS5700009	VS5700009	Malawi	*Homo sapiens*	Human	Blood serum	Smits et al., 2013
KF031465	Cyclovirus VN hcf1	hcf1	Vietnam	*Homo sapiens*	Human	Cerebrospinal fluid	Tan et al., 2013
KF031466	Cyclovirus VN hcf2	hcf2	Vietnam	*Homo sapiens*	Human	Cerebrospinal fluid	Tan et al., 2013
KF031467	Cyclovirus VN hcf3	hcf3	Vietnam	*Homo sapiens*	Human	Cerebrospinal fluid	Tan et al., 2013
KF031468	Cyclovirus VN hcf4	hcf4	Vietnam	*Homo sapiens*	Human	Cerebrospinal fluid	Tan et al., 2013
KF031469	Cyclovirus VN hcf5	hcf5	Vietnam	*Homo sapiens*	Human	Cerebrospinal fluid	Tan et al., 2013
KF031470	Cyclovirus VN ps1	ps1	Vietnam	*Sus scrofa*	Wild pig	Faeces	Tan et al., 2013

Genbank Accession	Cycloviruses description	Acronyms	Country	Isolation source	Common name	Sample type	References
KF031471	Cyclovirus VN cs1	cs1	Vietnam	*Gallus gallus*	Chicken	Faeces	Tan et al., 2013
KF726984	Human cyclovirus 7078A	7078A	Chile	*Homo sapiens*	Human	Respiratory secretion	Phan et al., 2014
KF726985	Human cyclovirus 7081A	7081A	Chile	*Homo sapiens*	Human	Respiratory secretion	Phan et al., 2014
KF726986	Human cyclovirus 5841A	5841A	Chile	*Homo sapiens*	Human	Respiratory secretion	Phan et al., 2014
KF726987	Human cyclovirus 7046A	7046A	Chile	*Homo sapiens*	Human	Respiratory secretion	Phan et al., 2014
KJ641710	Bat circovirus*	BtMbly-CyV	China	*Myotis blythii*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641712	Bat circovirus*	BtRp-CyV-3	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641714	Bat circovirus*	BtRp-CyV-14	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641715	Bat circovirus*	BtRp-CyV-52	China	*Rhinolophus pusillus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641717	Bat circovirus*	BtMspp.-CyV	China	*Myotis* spp.	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641720	Bat circovirus*	BtTp-CyV-2	China	*Tylonycteris pachypus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641728	Bat circovirus*	BtPa-CV-2/	China	*Plecotus auritus*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641734	Bat circovirus*	BtVS-CyV	China	*Vespertilio superans*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ641740	Bat circovirus*	BtRf-CyV-24	China	*Rhinolophus ferrumequinum*	Insects	Pharyngeal & rectal swabs	Wu et al.,2015
KJ831064	Cyclovirus SL-108277	SL_108277	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	Phan et al., 2015
KM017740	Feline cyclovirus	Feline	USA	*Felis catus*	Cat	Faeces	Zhang et al., 2014
KM382269	Bat circovirus POA/2012/II	BatCV_POA_2012_II	Southern Brazil	*Molossus molossus, Tadarida brasiliensis*	Bat	Faeces	Lima et al., 2015
KM382270	Bat circovirus POA/2012/VI	BatCV_POA_2012_VI	Southern Brazil	*Molossus molossus, Tadarida brasiliensis*	Bat	Faeces	Lima et al., 2015
KM392284	Swine cyclovirus SC_CGS96	SC_CGS96	Cameroon	*Sus scrofa*	Wild pig	Faeces	Garigliany et al., 2014
KM392285	Swine cyclovirus SC_CGS88	SC_CGS88	Cameroon	*Sus scrofa*	Wild pig	Faeces	Garigliany et al., 2014
KM392286	Swine cyclovirus SC_CGS77	SC_CGS77	Cameroon	*Sus scrofa*	Wild pig	Faeces	Garigliany et al., 2014
KM392287	Human cyclovirus VN-like HC_CGS288	HC_CGS288	Madagascar	*Homo sapiens*	Human	Faeces	Garigliany et al., 2014
KM392288	Human cyclovirus VN-like HC_CGS202	HC_CGS202	Madagascar	*Homo sapiens*	Human	Faeces	Garigliany et al., 2014
KM392289	Human cyclovirus VN-like HC_CGS104	HC_CGS104	Madagascar	*Homo sapiens*	Human	Faeces	Garigliany et al., 2014
KP151567	Cyclovirus NI-204	NI_204	Nicaragua	*Homo sapiens*	Human	Faeces	Phan et al., 2015
KR902499	Cyclovirus Equ1	Equ1	USA	*Equus caballus*	Horse	Nasal secretions	Li et al., 2015
LC018134	Cyclovirus TsCyV-1	TsCyV-1	Japan	*Callosciurus erythraeus thaiwanensis*	Taiwan squirrels	Stomach contents	Sato et al., 2015
KT732785	**Pacific flying fox associated cyclovirus-1**	**PfffaCyV-1**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**
KT732786	**Pacific flying fox associated cyclovirus-2**	**PfffaCyV-2**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**
KT732787	**Pacific flying fox associated cyclovirus-3**	**PfffaCyV-3**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**
KT732788	**Pacific flying fox associated cyclovirus-3**	**PfffaCyV-3**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**
KT732789	**Pacific flying fox associated cyclovirus-3**	**PfffaCyV-3**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**

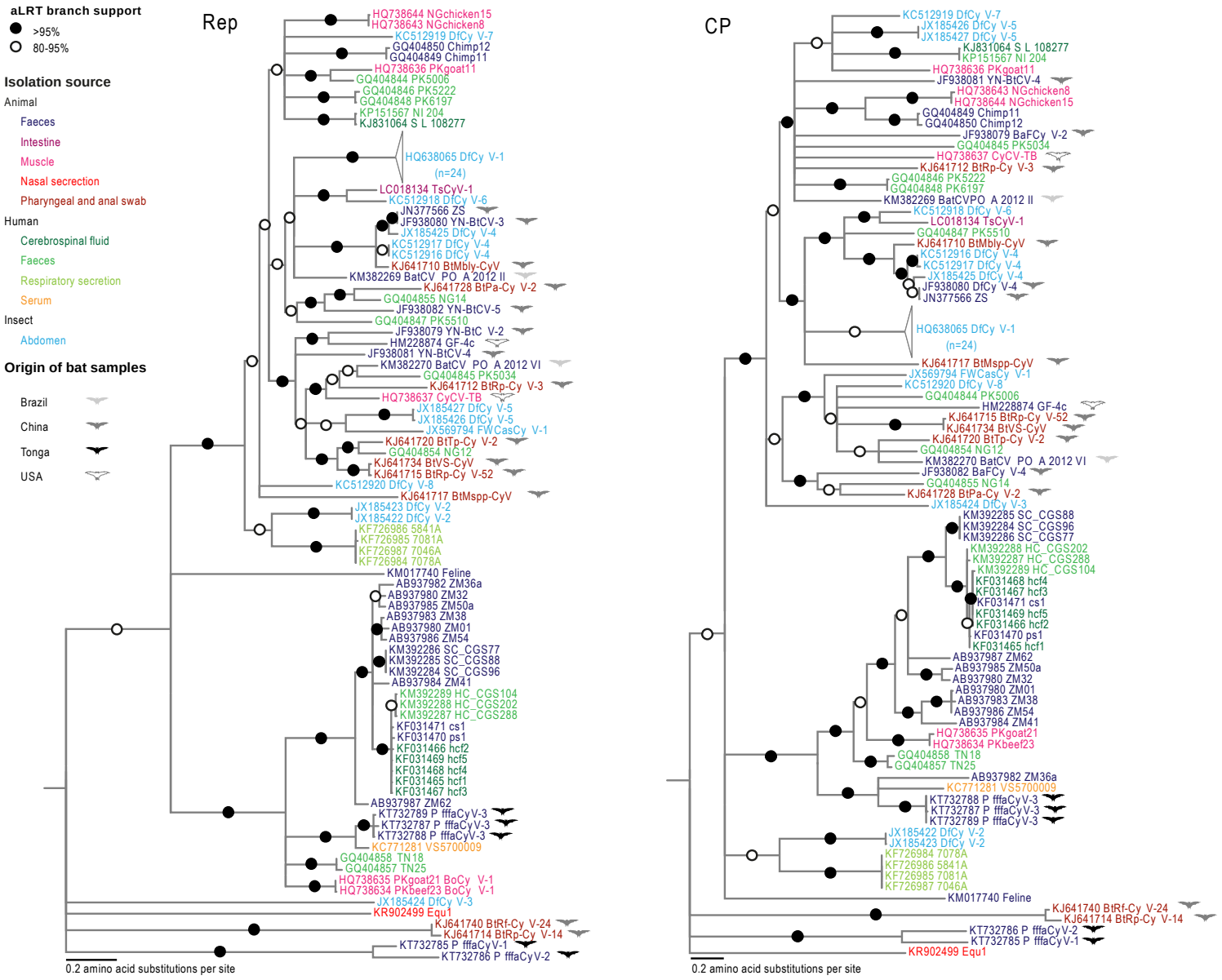


Figure 2.4: See next page for figure legend

Figure 2.4: Maximum-likelihood phylogenetic trees of the Rep and CP amino acid sequences of cycloviruses recovered from this study together with those available in GenBank. Branches with <80% aLRT support have been collapsed. Coloured accession numbers and viral names represent the isolation source and the coloured bat cartoons represent global sampling locations.

Table 2.5: Conserved motifs identified in the Reps of cycloviruses, gemycircularviruses, unclassified CRESS DNA viruses and the multi-component virus identified in this study.

				RCR motifs				SH3 helicase motifs		
Viral grouping	**Accession #**	**Sequence ID**	**Genome type***	**I**	**II**	**III**	**GRS motif**	**Walker-A**	**Walker-B**	**Motif C**
Cycloviruses	KT732785	PfffaCyV-1	II	VFTLNN	KHLQG	YLPH		GAPGVGKS	IIDDF	ITSN
	KT732786	PfffaCyV-2	II	VFTLNN	KHLQG	YLPH		GTPGVGKS	IIDDY	ITTN
	KT732787	PfffaCyV-3	II	CWTLNN	KHLQG	YCSK		GATGLGKS	VVIDEF	ITSN
	KT732788	PfffaCyV-3	II	CWTLNN	KHLQG	YCSK		GATGLGKS	VIDDF	ITSN
	KT732789	PfffaCyV-3	II	CWTLNN	KHLQG	YCSK		GATGLGKS	VIDDF	ITSN
Gemycircularviruses	KT732790	PfffaGmV-1	II	LLTYSQ	THLHA	YAIK	RRFDVEGFHPNIQPCG	GETRLGKT	VLDDI	WLMN
	KT732791	PfffaGmV-1	II	LLTYSQ	THLHA	YAIK	RRFDVEGFHPNIQPCG	GETRLGKT	VLDDI	WLMN
	KT732792	PfffaGmV-2	II	LFTYSQ	THLHV	YAIK	KIFDCEGRHPNVSASR	GKSRTGKT	VFDDI	WLSN
	KT732793	PfffaGmV-2	II	LFTYSQ	THLHV	YAVK	KIFDCEGRHPNVSASR	GKSRTGKT	VFDDI	WLSN
	KT732794	PfffaGmV-3	II	LLTYAQ	THLHV	YATK	DYFDVEGHHPNIVPSR	GPSRLGKT	VFDDM	WLAN
	KT732795	PfffaGmV-4	II	LLTYPQ	THLHA	YAIK	DFFDVGGHHPNIAPSR	GPSRLGKT	VFDDM	WLAN
	KT732796	PfffaGmV-4	II	LLTYPQ	THLHA	YAIK	DFFDVGGHHPNIAPSR	GPSRLGKT	VFDDM	WLAN
	KT732797	PfffaGmV-5	II	LVTYPQ	LHLHV	YAIK	NIFDVDGRHPNRAPSK	GGTRTGKT	VFDDI	WVCN
	KT732798	PfffaGmV-6	II	MLTYPT	PHIHV	YVAK	ATFKIGTRVPNIRVRR	GATRLGKT	IFDDM	FICN
	KT732799	PfffaGmV-6	II	MLTYPT	PHIHV	YVAK	ATFKIGTRVPNIRVRR	GATRLGKT	IFDDM	FICN
	KT732800	PfffaGmV-7	II	MVTFVR	PHYHA	YVGK	KTFQVAGRSPNIRVRR	GGSRFGKT	VFNDM	WTCN
	KT732801	PfffaGmV-8	II	LFTYSQ	IHFHV	YAIK	RVFDVGGKHPNIKPIG	GPYGCGKT	IFDDW	WLCN
	KT732802	PfffaGmV-8	II	LFTYAQ	IHYHV	YAIK	RIFDVGGKHPNIKPIG	GPYGCGKT	IFDDW	WLCN
	KT732803	PfffaGmV-9	II	LFTYSQ	IHFHV	YAIK	RVFDVGGKHPNIQPIG	GPYGCGKT	IFDDW	WLCN
	KT732804	PfffaGmV-10	II	LLTYAH	FHFHV	YATK	DVFDVDGYHPNIEPSR	GPTRLGKT	IMDDI	WCYN
	KT732805	PfffaGmV-10	II	LLTYAH	FHFHV	YATK	DVFDVDGYHPNIEPSR	GPTRLGKT	IMDDI	WCSN
	KT732814	PfffaGmV-13	II	LLTYAQ	IHLHV	YAIK	GIFDVGGRHPNVVASW	GESRLGKT	VFDDM	WGSN
	KT732812	PfffaGmV-11	II	ILTYSQ	IHLHV	YAIK	DVFDVGGYPPNIAKCG	GDTRLGKT	VFDDI	WLSN
	KT732807	PfffaGmV-11	II	LITYSQ	IHLHV	YAIK	DVFDVGGCHPNIAKCG	GDTRLGKT	VFDDI	WLSN
	KT732809	PfffaGmV-11	II	LITYSQ	IHLHV	YAIK	DVFDVGGCHPNIAKCG	GDTRLGKT	VFDDI	WLSN
	KT732810	PfffaGmV-11	II	FITYSQ	IHLHV	YAIK	DVFDVGGCHPNIAKCG	GDTRLGKT	VFDDI	WLSN
	KT732808	PfffaGmV-11	II	LITYSQ	IHLHV	YAIK	DVFDVGGCHPNIAKCG	GDTRLGKT	VFDDI	WLSN
	KT732811	PfffaGmV-11	II	LLTYSQ	IHLHV	YAIK	DVFDVGGCHPNIAKCG	GDTRLGKT	VFDDI	WLSN
	KT732813	PfffaGmV-12	II	LLTYAQ	IHLHA	YAIK	RTFDVEGYHPNISPSR	GPSRMGKT	VFDDF	WLSN
	KT732806	PfffaGmV-14	II	LLTYSQ	LHLHV	YAIK	DVFDVGGHHPNIAKCG	GDTRLGKT	VFDDI	WLSN
Unclassified CRESS DNA viruses	KT732820	PfffaCV-1	V	CFTNFN	LHAQG	YCTK		GASGLGKS	LFDDF	ITSN
	KT732821	PfffaCV-1	V	CFTNFN	LHAQG	YCTK		GASGLGKS	LFDDF	ITSN
	KT732829	PfffaCV-2	I	LLTYPQ	EHVHV	YCRK		GPSGLGKS	DIDDL	FTSC
	KT732831	PfffaCV-2	I	LLTYPQ	EHVHV	YCRK		GPSGLGKS	DIDDL	FTSC

Viral grouping	Accession #	Sequence ID	Genome type*	RCR motifs			GRS motif	SH3 helicase motifs		
				I	II	III		Walker-A	Walker-B	Motif C
	KT732818	PfffaCV-3	II	VFTLNN	PHLQG	YCSK		-	IFDFS	CFAN
	KT732819	PfffaCV-4	I	CFTLNN	PHHQG	YCTK		GPPGTGKS	VIDEL	VTSN
	KT732822	PfffaCV-5	IV	VFTIFV	LHWQG	YCTK		GTPGTGKS	IVDDW	FTSN
	KT732823	PfffaCV-6	V	CFTAFA	KHIQG	YCKK		GPSFGIGKD	ISDFD	VTSN
	KT732824	PfffaCV-7	II	KFTHFK	-	-		GPPGTGKT	ILDEF	ICSN
	KT732825	PfffaCV-8	II	CFTVNN	KHLQGF	YCKK		GYPGSGKS	IIDDF	ITSN
	KT732784	PfffaCV-9	V	LITAHF	-	YCTK		GAPGVGKS	VIDDF	VTSN
	KT732827	PfffaCV-10	II	VWTSFK	LHWQG	YCQK		GTPGTGKS	IFDDF	FTSN
	KT732828	PfffaCV-11	V	CFTLNN	FHLQR	-		-	ISDDG	AGSN
	KT732830	PfffaCV-12	V	FLTYPH	LHIHA	YVKK		GPSGWGKT	ILDDL	LVGN
	KT732832	PfffaCV-13	V	CFTLNN	PHIQG	YCGK		GLTGTGKS	VLDDF	ITSN
	KT732833	PfffaCV-14	I	FLTYPQ	KHLHV	YVCK		GPPNVGKT	RWDDE	ILSN
	KT732834	PfffaCV-15	I	ILTFPQ	PHLHV	YVTK		GPRNLGKT	YSDDY	ILSN
Multicomponent CRESS DNA virus	KT732816	PfffaMCV-1 (DNA-R)	VI	CYTVNN	IHLQG	YCKK		GPTGTGKS	LIEDF	VTSN

* based Rosario et al. 2012b

2.4.3 Gemycircularviruses

From the Pacific flying fox faeces samples, I recovered 25 CRESS DNA viral sequences which are most closely related to gemycircularviruses. These 25 novel gemycircularviruses are grouped into 14 putative species (sharing <78% pairwise identity) as proposed in Kraberger *et al.* (2015a) and Sikorski *et al.* (2013b) and I have tentatively named as Pacific flying fox faeces-associated gemycircularviruses (PfffaGmV) 1 - 14 (PfffaGmV-1, n=2; PfffaGmV-2, n=2; PfffaGmV-3, n=1; PfffaGmV-4, n=2; PfffaGmV-5, n=1; PfffaGmV-6, n=2; PfffaGmV-7, n=1; PfffaGmV-8, n=2; PfffaGmV-9, n=1; PfffaGmV-10, n=2; PfffaGmV-11, n=6; PfffaGmV-12, n=1; PfffaGmV-13, n=1; PfffaGmV-14, n=1). The 25 PfffaGmVs encode putative functional Reps expressed from a putative spliced Rep transcript (Figure 2.2). A spliced Rep and a RepA are common features seen in some geminiviruses (Bernardo *et al.*, 2013; Varsani *et al.*, 2014b; Wright *et al.*, 1997), where each transcript is thought to be essential for replication and infection (Dekker *et al.*, 1991; Liu *et al.*, 1998; Wright *et al.*, 1997).

Among the gemycircularviruses there is ~48% diversity and the 14 PfffaGmV species from this study share <75% genome-wide pairwise identity to other gemycircularviruses. Two putative species of gemycircularviruses have previously been recovered from grasses (*Brachiaria deflexa* and *Saccharum hybrid*) and an adult dragonfly (*P. flavescens*) from Tonga (Male *et al.*, 2015; Rosario *et al.*, 2012a) (Figure 2.5; Table 2.6). These share low similarity to the PfffaGmVs from this study, sharing between 56-61% genome-wide identity. Genome-wide percentage pairwise identities of gemycircularviruses generated using SDT v1.2 (Muhire *et al.*, 2014) are provided in Supplementary Data 2.2.

The Rep of PfffaGmV-1 shares 79-84% pairwise identity to those of sewage associated gemycircularvirus-5 (KJ547635), human genital associated circular DNA virus-1 (KJ413144) and Meles meles faecal virus (JN704610) and the Rep of PfffaGmV-5 shares 83% pairwise identity with Hypericum japonicum associated circular DNA virus (KF413620). The Reps of PfffaGmVs, except PfffaGmV-1, recovered from this study share low percentage identity with those of other known gemycircularvirus Reps. The CPs of gemycircularviruses are overall more diverse than the Reps (Figure 2.5). Apart from the 25 gemycircularviruses identified in this study (Figure 2.5; Table 2.6), the only other gemycircularviruses recovered from bat so far are from pharyngeal and rectal swabs of an insectivorous bat (*Rhinolophus ferrumequinum*) from China (Wu *et al.*, 2015). A number of the Reps of gemycircularviruses

from bats cluster together. The Reps of PfffaGmV-2 and -5 cluster in a well-supported clade which contains sequences from a number of different sources: animals, plants, fungi, river and sewage. All Reps of PfffaGmVs have the GRS domain in addition to the conserved RCR and SF3 motifs. These are similar to those found in other gemycircularviruses (Table 2.5).

It is evident that gemycircularviruses are highly prevalent in nature and have been recovered from environmental, animal, insect and human samples from Brazil, Canada, China, Germany, Ghana, Nepal, Netherlands, New Zealand, Portugal, South Africa, Sri Lanka, Tonga, USA and Vietnam (Table 2.6). However, other than SsHADV-1 which confers hypovirulence in *S. sclerotiorum* (Yu *et al.*, 2010), it is unknown whether gemycircularviruses are pathogenic. Furthermore, for all but one gemycircularvirus the hosts are unknown but it is thought that they are probably associated with fungi based on Rep-like sequences that have been identified in fungal genomes (Liu *et al.*, 2011; Yu *et al.*, 2010).

Table 2.6: Summary of all known gemycircularviruses as of 1st October 2015.

Genbank Accession #	Gemycircularviruses description	Acronyms	Country	Isolation source	Common Name	Sample type	Reference
GQ365709	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	China	*Sclerotinia sclerotiorum*	*Sclerotinia sclerotiorum*	Mycelial samples	Yu et al., 2010
HQ335086	Mosquito VEM virus SDBVL G	MVemV	USA	*Culex erythrothorax*	Mosquito	Mosquito samples	Ng et al., 2011
JN704610	Meles meles fecal virus	MmFV	Netherlands	*Meles meles*	European badger	Rectal swab	van den Brand et al., 2012
JQ412057	Cassava associated circular DNA virus	CasCV	Ghana	*Manihot esculenta*	Cassava	Faeces	Dayaram et al., 2012
JX185428	Dragonfly-associated circular virus 3	DfasCV-3	Tonga	*Pantala flavescens*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185429	Dragonfly-associated circular virus 2	DfasCV-2	USA	*Erythemis simplicicollis*	Dragonfly	Abdomen	Rosario et al., 2012a
JX185430	Dragonfly-associated circular virus 1	DfasCV-1	USA	*Miathyria marcella*	Dragonfly	Abdomen	Rosario et al., 2012a
KF268025	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	Kraberger et al., 2013
KF268026	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	Kraberger et al., 2013
KF268027	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	Kraberger et al., 2013
KF268028	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	New Zealand	River Sediments	-	River Sediments	Kraberger et al., 2013
KF371630	Faecal-associated gemycircularvirus 12	FaGmV-12	New Zealand	*Struthio camelus*	Ostrich	Faeces	Sikorski et al., 2013b
KF371631	Faecal-associated gemycircularvirus 11	FaGmV-11	New Zealand	*Oryctolagus cuniculus*	Rabbit	Faeces	Sikorski et al., 2013b
KF371632	Faecal-associated gemycircularvirus 10	FaGmV-10	New Zealand	*Sturnus vulgaris*	European starling	Faeces	Sikorski et al., 2013b
KF371633	Faecal-associated gemycircularvirus 9	FaGmV-9	New Zealand	*Turdus merula*	Blackbird	Faeces	Sikorski et al., 2013b
KF371634	Faecal-associated gemycircularvirus 8	FaGmV-8	New Zealand	*Petroica traversi*	Chatham Island black robin	Faeces	Sikorski et al., 2013b
KF371635	Faecal-associated gemycircularvirus 7	FaGmV-7	New Zealand	*Anas platyrhynchos*	Mallard duck	Faeces	Sikorski et al., 2013b
KF371636	Faecal-associated gemycircularvirus 6	FaGmV-6	New Zealand	*Gerygone albofrontata*	Chatham Island warbler	Faeces	Sikorski et al., 2013b
KF371637	Faecal-associated gemycircularvirus 5	FaGmV-5	New Zealand	*Gerygone albofrontata*	Chatham Island warbler	Faeces	Sikorski et al., 2013b
KF371638	Faecal-associated gemycircularvirus 4	FaGmV-4	New Zealand	*Arctocephalus forsteri*	New Zealand fur seal	Faeces	Sikorski et al., 2013b
KF371639	Faecal-associated gemycircularvirus 3	FaGmV-3	New Zealand	*Gerygone albofrontata*	Chatham Island warbler	Faeces	Sikorski et al., 2013b
KF371640	Faecal-associated gemycircularvirus 2	FaGmV-2	New Zealand	*Sus scrofa*	Domestic pig	Faeces	Sikorski et al., 2013b
KF371641	Faecal-associated gemycircularvirus 1c	FaGmV-1c	New Zealand	*Turdus merula*	Blackbird	Faeces	Sikorski et al., 2013b
KF371642	Faecal-associated gemycircularvirus 1b	FaGmV-1b	New Zealand	*Turdus merula*	Blackbird	Faeces	Sikorski et al., 2013b
KF371643	Faecal-associated gemycircularvirus 1a	FaGmV-1a	New Zealand	*Ovis aries*	Sheep	Faeces	Sikorski et al., 2013b
KF413620	Hypericum japonicum associated circular DNA virus	HJasCV	Viet Nam	*Hypericum japonicum*	Hypericum	Leaf	Du et al., 2014
KJ413144	Human genital-associated circular DNA virus-1	HuGaGmC349	South Africa	*Homo sapiens*	Human	Cervical sample	unpublished
KJ547634	Sewage-associated gemycircularvirus-4	SaGmV-4	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547635	Sewage-associated gemycircularvirus-5	SaGmV-5	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547636	Sewage-associated gemycircularvirus-6	SaGmV-6	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547637	Sewage-associated gemycircularvirus-7a	SaGmV-7a	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547638	Sewage-associated gemycircularvirus-8	SaGmV-8	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547639	Sewage-associated gemycircularvirus-9	SaGmV-9	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547640	Sewage-associated gemycircularvirus-7b	SaGmV-7b	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547641	Sewage-associated gemycircularvirus-11	SaGmV-11	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a

Genbank Accession #	Gemycircularviruses description	Acronyms	Country	Isolation source	Common Name	Sample type	Reference
KJ547642	Sewage-associated gemycircularvirus-2	SaGmV-2	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547643	Sewage-associated gemycircularvirus-3	SaGmV-3	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547644	Sewage-associated gemycircularvirus-10a	SaGmV-10a	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ547645	Sewage-associated gemycircularvirus-10b	SaGmV-10b	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KJ641719	Bat gemycircularvirus 23 GD2012	BtMf-CV-23 GD2012	China	*Miniopterus fuliginosus*	Bat	Pharyngeal & rectal swabs	Wu et al., 2015
KJ641726	Bat gemycircularvirus 8 NM2013	BtRf-CV-8 NM2013	China	*Rhinolophus ferrumequinum*	Bat	Pharyngeal & rectal swabs	Wu et al., 2015
KJ641737	Bat gemycircularvirus Tibet2013	BtRh-CV-6 Tibet2013	China	*Rhinolophus hipposideros*	Bat	Pharyngeal & rectal swabs	Wu et al., 2015
KJ938717	Caribou feces-associated gemycircularvirus	FaGmV-13	Canada	*Rangifer tarandus*	Caribou	Faeces	Ng et al., 2014
KM510192	Bromus-associated circular DNA virus 3	BasCV-3	New Zealand	*Bromus hordeaceus*	Soft brome / Bull grass	Leaf	Kraberger et al., 2015b
KM598382	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	USA	*Ischnura ramburii*	Damselfly	Abdomen	Dayaram et al., 2015b
KM598383	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	USA	*Erythemis simplicicollis*	Dragonfly	Abdomen	Dayaram et al., 2015b
KM598384	Sclerotinia sclerotiorum hypovirulence associated DNA virus 1	SsHADV-1	USA	*Pantala hymenaea*	Dragonfly	Abdomen	Dayaram et al., 2015b
KM598385	Odonata associated gemycircularvirus-1	OdaGmV-1	USA	*Ischnura posita*	Damselfly	Abdomen	Dayaram et al., 2015b
KM598386	Odonata associated gemycircularvirus-1	OdaGmV-1	USA	*Pantala hymenaea*	Dragonfly	Abdomen	Dayaram et al., 2015b
KM598387	Odonata associated gemycircularvirus-2	OdaGmV-2	USA	*Aeshna multicolor*	Dragonfly	Abdomen	Dayaram et al., 2015b
KM598388	Odonata associated gemycircularvirus-2	OdaGmV-2	USA	*Libellula saturata*	Dragonfly	Abdomen	Dayaram et al., 2015b
KM821747	Sewage-associated gemycircularvirus-1	SaGmV-1	New Zealand	Sewage oxidation pond	-	Swage	Kraberger et al., 2015a
KP133075	Gemycircularvirus SL1	GemyCV-SL1	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	Phan et al., 2015
KP133076	Gemycircularvirus SL2	GemyCV-SL2	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	Phan et al., 2015
KP133077	Gemycircularvirus SL3	GemyCV-SL3	Sri Lanka	*Homo sapiens*	Human	Cerebrospinal fluid	Phan et al., 2015
KP133078	Gemycircularvirus BZ1	GemyCV-BZ1	Brazil	*Homo sapiens*	Human	Faeces	Phan et al., 2015
KP133079	Gemycircularvirus BZ2	GemyCV-BZ2	Brazil	*Homo sapiens*	Human	Faeces	Phan et al., 2015
KP133080	Gemycircularvirus NP	GemyCV-NP	Nepal	Untreated sewage	-	Sewage	Phan et al., 2015
KP263543	Badger faeces-associated gemycircularvirus	BafaGM588	Portugal	*Meles meles*	European badger	Faeces	Conceicao-Neto et al., 2015
KP263544	Mongoose feces-associated gemycircularvirus a	MoFaGmV181a	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	Conceicao-Neto et al., 2015
KP263545	Mongoose feces-associated gemycircularvirus b	MoFaGmV160b	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	Conceicao-Neto et al., 2015
KP263546	Mongoose feces-associated gemycircularvirus c	MoFaGmV541c	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	Conceicao-Neto et al., 2015
KP263547	Mongoose feces-associated gemycircularvirus d	MoFaGmV478d	Portugal	*Herpestes ichneumon*	Egyptian mongoose	Faeces	Conceicao-Neto et al., 2015
KR912221	Gemycircularvirus gemy-ch-rat1	Gemy-ch-rat1	China	*Rattus norvegicus*	Rat	Blood	Li et al., 2015b
KT253577	Poaceae associated gemycircularvirus-1	PaGmV-1	Tonga	*Brachiaria deflexa*	Signalgrass	Leaf	Male et al., 2015
KT253578	Poaceae associated gemycircularvirus-1	PaGmV-1	Tonga	*Brachiaria deflexa*	Signalgrass	Leaf	Male et al., 2015
KT253579	Poaceae associated gemycircularvirus-1	PaGmV-1	Tonga	*Saccharum hybrid*	Sugarcane	Leaf	Male et al., 2015
KT309029	Poecile atricapillus GI tract-associated gemycircularvirus	Gitract	USA	*Poecile atricapillus*	Black-capped chickadee	Buccal and cloacal swab	Hanna et al., 2015
LK931483	HCBI8.215 virus	HCBI8_215	Germany	*Bos taurus*	Cow	Healthy cow's blood serum	Lamberto et al., 2014
LK931484	HCBI9.212 virus	HCBI9_212	Germany	*Bos taurus*	Cow	Healthy cow's blood serum	Lamberto et al., 2014
LK931485	MSSI2.225 virus	MSSI2_225	Germany	*Homo sapiens*	Human	Multiple sclerosis patient's	Lamberto et al., 2014
KT732790	**Pacific flying fox faeces associated gemycircularvirus-1**	**PfffaGmV-1**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**
KT732791	**Pacific flying fox faeces associated gemycircularvirus-1**	**PfffaGmV-1**	**Tonga**	***Pteropus tonganus***	**Bat**	**Faeces**	**This study**

Genbank Accession #	Gemycircularviruses description	Acronyms	Country	Isolation source	Common Name	Sample type	Reference
KT732792	Pacific flying fox faeces associated gemycircularvirus-2	PfffaGmV-2	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732793	Pacific flying fox faeces associated gemycircularvirus-2	PfffaGmV-2	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732794	Pacific flying fox faeces associated gemycircularvirus-3	PfffaGmV-3	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732795	Pacific flying fox faeces associated gemycircularvirus-4	PfffaGmV-4	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732796	Pacific flying fox faeces associated gemycircularvirus-4	PfffaGmV-4	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732797	Pacific flying fox faeces associated gemycircularvirus-5	PfffaGmV-5	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732798	Pacific flying fox faeces associated gemycircularvirus-6	PfffaGmV-6	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732799	Pacific flying fox faeces associated gemycircularvirus-6	PfffaGmV-6	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732800	Pacific flying fox faeces associated gemycircularvirus-7	PfffaGmV-7	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732801	Pacific flying fox faeces associated gemycircularvirus-8	PfffaGmV-8	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732802	Pacific flying fox faeces associated gemycircularvirus-8	PfffaGmV-8	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732803	Pacific flying fox faeces associated gemycircularvirus-9	PfffaGmV-9	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732804	Pacific flying fox faeces associated gemycircularvirus-10	PfffaGmV-10	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732805	Pacific flying fox faeces associated gemycircularvirus-10	PfffaGmV-10	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732807	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732808	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732809	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732810	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732811	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732812	Pacific flying fox faeces associated gemycircularvirus-11	PfffaGmV-11	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732813	Pacific flying fox faeces associated gemycircularvirus-12	PfffaGmV-12	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732814	Pacific flying fox faeces associated gemycircularvirus-13	PfffaGmV-13	Tonga	*Pteropus tonganus*	Bat	Faeces	This study
KT732806	Pacific flying fox faeces associated gemycircularvirus-14	PfffaGmV-14	Tonga	*Pteropus tonganus*	Bat	Faeces	This study

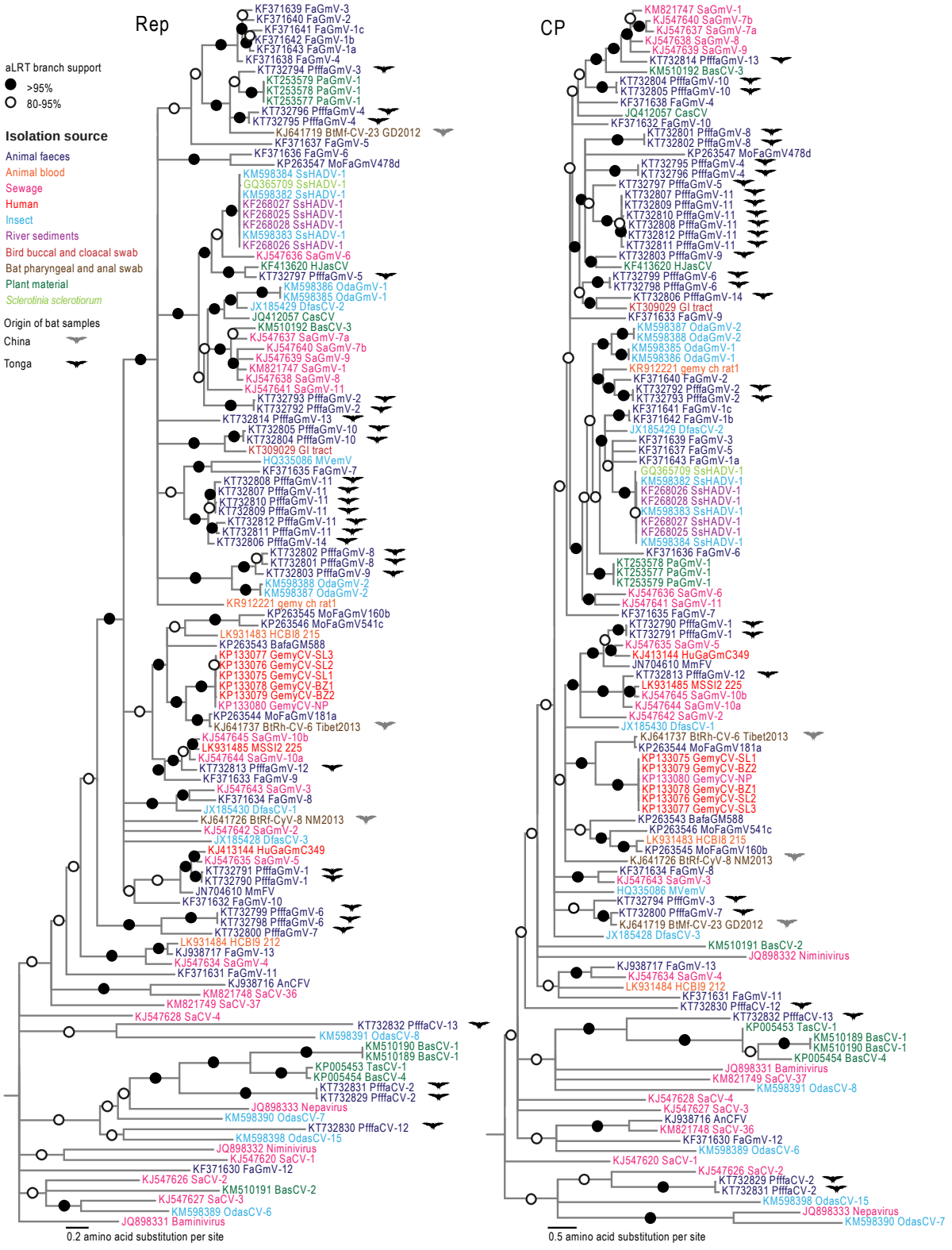


Figure 2.5: Maximum-likelihood phylogenetic trees of the Rep and CP of gemycircularviruses, and gemycircularvirus-like sequences recovered from this study together with those available in GenBank. Branches with <80% aLRT support have been collapsed. Coloured accession numbers and viral names represent the isolation source and the coloured bat cartoons represent the global sampling locations.

2.4.4 Unclassified CRESS DNA viruses

I identified 17 circular molecules which are putative novel CRESS DNA viruses, these share very low levels of sequence similarity to other known CRESS DNA viruses (Table 2.7). I grouped these into 15 species based on their Reps sharing less than 75% pairwise amino acid identity and I tentatively named these as Pacific flying fox faeces-associated circular DNA virus (PfffaCV) - 1 through 15. The CRESS DNA viral-like sequences range in size from 1757 nts to 2732 nts and all contain two large ORFs that encode a Rep and a putative CP (Figure 2.2). The putative CPs were identified based on similarities of these ORFs to known CRESS DNA viruses (Table 2.7) and the presence of basic residues on the N-terminus portion of the ORF (Rosario *et al.*, 2012c). PfffaCV-1, PfffaCV-6, PfffaCV-11, PfffaCV-12 and PfffaCV-15 contain unidirectional ORFs, whereas PfffaCV-2, PfffaCV-3, PfffaCV-4, PfffaCV-5, PfffaCV-7, PfffaCV-8, PfffaCV-9, PfffaCV-10, PfffaCV-13 and PfffaCV-14 have bidirectionally transcribed ORFs (Figure 2.2). Based on the genome type classification proposed by Rosario *et al.* (2012c), the 17 PfffaCVs represent four genome types namely type I (n=5), II (n=4), IV (n=1) and V (n=7) (Table 2.5).

Table 2.7: Summary of the BLASTp hits of the putative Rep, CP and unknown proteins encoded by the unclassified CRESS DNA viruses, putative multi-component and circular molecules recovered in this study.

Accession #	Query	ORFs	Virus	Accession #	% Pairwise Identity	E Value	Query coverage	Isolation source	Country
KT732820 &	PfffaCV-1	CP	-	-	-	-	-	-	-
KT732821		Rep	Bat circovirus BtRp-CV	KJ641731	33%	$2x10^{-45}$	84%	Bat pharyngeal and anal swab	China
KT732829 &	PfffaCV-2	CP	Avon-Heathcote Estuary associated circular virus-25	KM874355	27%	$9x10^{-13}$	50%	*Austrovenus stutchburyi*	New Zealand
KT732831		Rep	Odonata-associated circular virus-7	KM598390	34%	$5x10^{-45}$	85%	*Libellula quadrimaculata*	USA
KT732818	PfffaCV-3	CP	Cyanoramphus nest associated circular X DNA virus	JX908739	31%	$1x10^{-16}$	77%	*Cyanoramphus auriceps* nest	New Zealand
		Rep	Cyanoramphus nest associated circular X DNA virus	JX908739	44%	$2x10^{-93}$	88%	*Cyanoramphus auriceps* nest	New Zealand
KT732819	PfffaCV-4	CP	Uncultured marine virus	KR528558	39%	$4x10^{-03}$	29%	Saanich Inlet	Canada
		Rep	Dragonfly larvae associated circular virus-3	KF738876	44%	$1x10^{-76}$	97%	*Procordulia grayi*	New Zealand
KT732822	PfffaCV-5	CP	-	-	-	-	-	-	-
		Rep	Bat circovirus	KJ641733	33%	$1x10^{-21}$	97%	*Miniopterus fuliginosus*	China
KT732823	PfffaCV-6	CP	-	-	-	-	-	-	-
		Rep	Odonata-associated circular virus-13	KM598396	40%	$4x10^{-59}$	77%	*Libellula quadrimaculata*	USA
KT732824	PfffaCV-7	CP	-	-	-	-	-	-	-
		Rep	Uncultured marine virus	JX904523	33%	$1x10^{-29}$	85%	Saanich Inlet salt water	Canada
KT732825	PfffaCV-8	CP	-	-	-	-	-	-	-
		Rep	Silurus glanis circovirus	JQ011378	43%	$8x10^{-62}$	94%	*Silurus glanis*	Hungary
KT732784	PfffaCV-9	Unk1	-	-	-	-	-	-	-
		Unk2	-	-	-	-	-	-	-
		Rep	Bat circovirus BtMf-CV	KJ641733	39%	$9x10^{-43}$	74%	Bat pharyngeal and rectal swab	China
KT732827	PfffaCV-10	CP	Acheta domesticus volvovirus	KC543331	33%	$5x10^{-33}$	77%	*Acheta domesticus*	Canada
		Rep	Acheta domesticus volvovirus	KC543331	41%	$3x10^{-60}$	95%	*Acheta domesticus*	Canada
KT732828	PfffaCV-11	CP	Uncultured marine virus	JX904457	29%	$1x10^{-10}$	65%	Saanich Inlet	Canada
		Rep	Rodent stool-associated circular virus	JF755403	40%	$1x10^{-48}$	88%	*Microtus pennsylvanicus* faeces	USA
KT732830	PfffaCV-12	CP	-	-	-	-	-	-	-
		Rep	Avon-Heathcote Estuary associated circular virus 25	KM874357	29%	$7x10^{-31}$	90%	Benthic sediment	New Zealand
KT732832	PfffaCV-13	CP	-	-	-	-	-	-	-
		Rep	Bat circovirus BtTp-CV	KJ641721	38%	$6x10^{-45}$	81%	Bat pharyngeal and rectal swab	China
KT732833	PfffaCV-14	CP	Sewage-associated circular DNA virus-26	KM874359	31%	$8x10^{-16}$	58%	Sewage oxidation pond	New Zealand
		Rep	Sewage-associated circular DNA virus-18	KM821753	35%	$2x10^{-43}$	80%	Sewage oxidation pond	New Zealand
KT732834	PfffaCV-15	CP	Sewage-associated circular DNA virus-18	KM821753	39%	$5x10^{-39}$	81%	Sewage oxidation pond	New Zealand
		Rep	Sewage-associated circular DNA virus-18	KM821753	63%	$1x10^{-168}$	100%	Sewage oxidation pond	New Zealand
KT732815	PfffaMCV-1(DNA-S)	CP	Odonata-associated circular virus-11	KM598394	24%	$2x10^{-08}$	68%	*Erythemis simplicicollis*	USA
KT732816	PfffaMCV-1(DNA-R)	Rep	Circoviridae 19 LDMD-2013	KF133826	48%	$3x10^{-86}$	96%	Ocean water	USA
KT732817	PfffaMCV-1(DNA-U1)	Unk1	-	-	-	-	-	-	-
KT732783	PfffaCM-1	Rep *	*Xanthomonas axonopodis* replication protein	WP_017161479	69%%	0	98%	Unknown	Unknown
KT732826	PfffaCM-2	CP	Bat circovirus	KJ641733	40%	$2x10^{-45}$	73%	*Miniopterus fuliginosus*	China
		Unk1	-	-	-	-	-	-	-

* = Replication protein

The Reps and CPs encoded by PfffaCV-2, PfffaCV-12 and PfffaCV-13 are somewhat related to those of gemycircularviruses and geminiviruses and hence are included in the Rep and CP analyses with the gemycircularviruses (Figure 2.5). These gemycircularvirus-like sequences group with the Odonata-associated circular viruses (OdasCV-6, -7, -8 and -15; KM598389 - KM598391 and KM598398) (Dayaram *et al.*, 2015b), Trifolium-associated circular DNA virus-1 (TasCV-1; KP005453) and Bromus-associated circular DNA viruses (BasCV-1, -2 and -4; KM510189 - KM510191 and KP005454) (Kraberger *et al.*, 2015b), Sewage-associated circular DNA viruses (SaCV-1, -2, -3 and -4; KJ547620 and KJ547626 - KJ547628) (Kraberger *et al.*, 2015a) and three other sewage-associated viruses (Baminivirus, Niminivirus and Nepavirus; JQ898331 - JQ898333) (Ng *et al.*, 2012) (Figure 2.5). However, their genomes are slightly larger (~2500-2600 nts) compared to those of gemycircularviruses (~2000-2300 nts). A summary of the BLASTp analysis of the putative Rep and CPs of the 15 novel species of unclassified CRESS DNA viruses are provided in Table 2.7.

The conserved residues in most motifs of PfffaCV-1, -3, -4 and -8 Reps are similar to other circoviruses and cycloviruses (Table 2.5). Reps of PfffaCV -4, -5, -6, -8, -10, -14 and -15 contain all the conserved motifs (Table 2.5) similar to other CRESS DNA viral families reviewed in Rosario *et al.* (2012c). However, motif II was not identified in PfffaCV-7 and -9, motif III was missing in PfffaCV-7 and -11, also PfffaCV-3 and -11 have no Walker-A motif (Table 2.5). Interestingly, Reps of PfffaCV-2, -12 and -13 are gemycircularvirus-like (Figure 2.5) but they do not contain the GRS domain which is present in the Reps of other gemycircularviruses, however, PfffaCV-2 and -12 contain motif I which is similar to Reps of some gemycircularviruses (Table 2.5).

2.4.5 Putative multi-component virus

Three ssDNA molecules of 1143-1163 nts, each encoding a single ORF were recovered. One ORF encodes a Rep, the second a putative CP and the third ORF had no similarities with any sequences in GenBank. Interestingly, all three of these molecules have a 58 nt region which is 100% identical, labelled the common region stemloop (CRSL) (Figure 2.3). Additionally, the Rep and CP encoding molecules have a 162 nt common region, labelled common region major that is 100% identical (Figure 2.3). The CRSL (58 nt region) contains a highly conserved nonanucleotide motif (TAGTATTAC) across all components (Figure 2.3). Based on the knowledge of multi-component viruses in the *Nanoviridae* family and bipartite begomoviruses of the *Geminiviridae* family, a Rep encoded by one molecule can initiate

replication of cognate molecules by binding to their iterons and nicking the stemloop at the nonanucleotide motif, I postulate that these three components are part of a novel multi-component virus. I have putatively named this novel multi-component virus as Pacific flying fox faeces-associated multi-component virus-1 (PfffaMCV-1). The PfffaMCV-1 Rep shares 48% pairwise identity with the Rep of circoviridae 19 LDMD-2013 (KF133826) recovered from USA (McDaniel *et al.*, 2014) (Table 2.7). The PfffaMCV-1 CP shares 24% pairwise identity with Odonata-associated circular virus-11 (KM598394) isolated from USA (Dayaram *et al.*, 2015b) (Table 2.7). Following the nomenclature used for nanoviruses the three DNA components of PfffaMCV-1 have been named PfffaMCV-1 (DNA-R; Rep encoding), PfffaMCV-1 (DNA-S; CP encoding) and PfffaMCV-1 (DNA-U1; unknown ORF). I note that the common region major of PfffaMCV-1 (DNA-R and DNA-S) shares no similarities with the members of the *Nanoviridae* family. On the other hand, the nonanucleotide motif TAGTATTAC is similar to that of members of the *Nanovirus* genus of the *Nanoviridae* family and their associated Rep-encoding satellite molecules, and begomovirus-associated alphasatellites (Bell *et al.*, 2002; Briddon *et al.*, 2004; Briddon & Stanley, 2006; Horser *et al.*, 2001; Katul *et al.*, 1998; Mansoor *et al.*, 1999; Saunders & Stanley, 1999; Wu *et al.*, 1994; Zhou, 2013).

The Rep of the putative multi-component virus (PfffaMCV-1) contains the conserved residues seen in the Reps motifs of eukaryotic CRESS DNA viruses. The RCR motif I is similar to that found in begomovirus-associated alphasatellites, the RCR motif II similar to those in begomovirus alphasatellites, circoviruses and cycloviruses, the RCR motif III is similar to those in begomovirus-associated alphasatellites, circoviruses, cycloviruses, geminiviruses and nanoviruses (Table 2.5). The Walker A, B and motif C are similar to those found in Reps of circoviruses and cycloviruses (Rosario *et al.*, 2012c).

2.4.6 Circular DNA molecules

Two further novel circular DNA molecules were identified, both of which do not exhibit a Rep similar to that encoded by eukaryotic CRESS DNA viruses. I tentatively named these Pacific flying foxes faeces-associated circular molecule (PfffaCM)-1 and -2 each containing a genome size of 1957 nt and 2255 nt, respectively (Figure 2.2). BLASTx analysis shows that the ORF of PfffaCM-1 shares 69% pairwise identity (98% coverage) with a *Xanthomonas axonopodis* replication protein (WP_017161479; Table 2.7) that is associated with plasmid replication (pfam02486). Therefore it is highly likely that this molecule is associated with

prokaryotes and could possibly be a non-viral circular DNA mobile element. One of the large ORFs of PfffaCM-2 shares 40% (73% coverage) pairwise identity with the CP of a Bat circovirus (KJ641733) (Table 2.7) and hence this could be a subgenomic molecule or a component of a multi-component virus.

2.5 Concluding remarks

Viral metagenomic studies have shown that animal faecal matter contains a high diversity of viruses and thus can be used to explore viral diversity in the environment. Through this viral metagenomics approach study of Pacific flying fox faeces from four roosting sites in Tonga in 2014 and 2015, I identified five cycloviruses, 25 gemycircularviruses, 17 unclassified CRESS DNA viruses, two circular DNA molecules and a putative novel multi-component virus with three cognate molecules.

A number of viruses were identified in more than one sampling site in Tonga suggesting these viruses have a broad distribution across the island amongst the Pacific flying fox colonies. Several species were identified in both 2014 and 2015 suggesting these viruses are persistently associated with faecal matter of Pacific flying foxes. The presence of PfffaCyV isolates in a well-supported clade which also contains isolates from human and domestic animals suggests that these viruses may also be circulating in animals other than bats in Tonga. PfffaCyV-3s share high similarity (~77% genome-wide identity) with Human cyclovirus VS5700009 (KC771281) which was identified in patient with paraplegia of unknown aetiology in Malawi (Smits *et al.*, 2013). Taking into consideration the observations by (Garigliany *et al.*, 2014) and (Tan *et al.*, 2013) where they found cyclovirus CyCN-VN in different geographical locations (Africa and Asia) and in humans and domestic animals, it is possible that the cycloviruses identified in Pacific flying foxes, and more so PfffaCyV-3 which is most closely related to CyCN-VN, may be found associated with other animals in Tonga.

In the Tongan archipelago, prior to this study only six CRESS-DNA virus species had been previously identified. The findings from this study contribute significantly to the knowledge of viruses circulating in Tonga and in supporting the current view that the diversity of CRESS DNA viruses is grossly underestimated.

2.6 References

Abascal, F., Zardoya, R. & Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Banack, S. A. (1998). Diet selection and resource use by flying foxes (Genus Pteropus). *Ecology* **79**, 1949-1967.

Basso, M. F., da Silva, J. C., Fajardo, T. V., Fontes, E. P. & Zerbini, F. M. (2015). A novel, highly divergent ssDNA virus identified in Brazil infecting apple, pear and grapevine. *Virus Res* **210**, 27-33.

Bell, K. E., Dale, J. L., Ha, C. V., Vu, M. T. & Revill, P. A. (2002). Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Archives of virology* **147**, 695-707.

Bernardo, P., Golden, M., Akram, M., Naimuddin, Nadarajan, N., Fernandez, E., Granier, M., Rebelo, A. G., Peterschmitt, M., Martin, D. P. & Roumagnac, P. (2013). Identification and characterisation of a highly divergent geminivirus: evolutionary and taxonomic implications. *Virus Res* **177**, 35-45.

Blinkova, O., Victoria, J., Li, Y., Keele, B. F., Sanz, C., Ndjango, J. B., Peeters, M., Travis, D., Lonsdorf, E. V., Wilson, M. L., Pusey, A. E., Hahn, B. H. & Delwart, E. L. (2010). Novel circular DNA viruses in stool samples of wild-living chimpanzees. *J Gen Virol* **91**, 74-86.

Breitbart, M., Benner, B. E., Jernigan, P. E., Rosario, K., Birsa, L. M., Harbeitner, R., Fulford, S., Graham, C., Walters, A., Goldsmith, D. B., Berger, S. A. & Nejstgaard, J. C. (2015). Discovery, prevalence, and persistence of novel circular single-stranded DNA viruses in the ctenophores Mnemiopsis leidyi and Beroe ovata. *Front Microbiol* **6**, 1426.

Briddon, R. W., Bull, S. E., Amin, I., Mansoor, S., Bedford, I. D., Rishi, N., Siwatch, S. S., Zafar, Y., Abdel-Salam, A. M. & Markham, P. G. (2004). Diversity of DNA 1: a satellite-like molecule associated with monopartite begomovirus-DNA beta complexes. *Virology* **324**, 462-474.

Briddon, R. W. & Stanley, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology* **344**, 198-210.

Brook, C. E. & Dobson, A. P. (2015). Bats as 'special' reservoirs for emerging zoonotic pathogens. *Trends Microbiol* **23**, 172-180.

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. (2006). Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev* **19**, 531-545.

Castrignano, S. B., Nagasse-Sugahara, T. K., Kisielius, J. J., Ueda-Ito, M., Brandao, P. E. & Curti, S. P. (2013). Two novel circo-like viruses detected in human feces:

complete genome sequencing and electron microscopy analysis. *Virus Res* **178**, 364-373.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. L. & Pogranichniy, R. M. (2014a). Identification of a novel single-stranded circular DNA virus in pig feces. *Genome announcements* **2**, 00347-00314.

Cheung, A. K., Ng, T. F., Lager, K. M., Alt, D. P., Delwart, E. L. & Pogranichniy, R. M. (2014b). Unique circovirus-like genome detected in pig feces. *Genome announcements* **2**, 00251-00214.

Cheung, A. K., Ng, T. F., Lager, K. M., Bayles, D. O., Alt, D. P., Delwart, E. L., Pogranichniy, R. M. & Kehrli, M. E. (2013). A divergent clade of circular single-stranded DNA viruses from pig feces. *Archives of virology* **158**, 2157-2162.

Cheung, A. K., Ng, T. F. F., Lager, K. M., Alt, D. P., Delwart, E. & Pogranichniy, R. M. (2015). Identification of several clades of novel single-stranded circular DNA viruses with conserved stem-loop structures in pig feces. *Archives of virology* **160**, 353-358.

Conceicao-Neto, N., Zeller, M., Heylen, E., Lefrere, H., Mesquita, J. R. & Matthijnssens, J. (2015). Fecal virome analysis of three carnivores reveals a novel nodavirus and multiple gemycircularviruses. *Virology Journal* **12**, e79.

Dayaram, A., Galatowitsch, M., Harding, J. S., Arguello-Astorga, G. R. & Varsani, A. (2014). Novel circular DNA viruses identified in Procordulia grayi and Xanthocnemis zealandica larvae using metagenomic approaches. *Infection Genetics and Evolution* **22**, 134-141.

Dayaram, A., Goldstien, S., Arguello-Astorga, G. R., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2015a). Diverse small circular DNA viruses circulating amongst estuarine molluscs. *Infection Genetics and Evolution* **31**, 284-295.

Dayaram, A., Opong, A., Jaschke, A., Hadfield, J., Baschiera, M., Dobson, R. C., Offei, S. K., Shepherd, D. N., Martin, D. P. & Varsani, A. (2012). Molecular characterisation of a novel cassava associated circular ssDNA virus. *Virus Res* **166**, 130-135.

Dayaram, A., Potter, K. A., Moline, A. B., Rosenstein, D. D., Marinov, M., Thomas, J. E., Breitbart, M., Rosario, K., Arguello-Astorga, G. R. & Varsani, A. (2013). High global diversity of cycloviruses amongst dragonflies. *J Gen Virol* **94**, 1827-1840.

Dayaram, A., Potter, K. A., Pailes, R., Marinov, M., Rosenstein, D. D. & Varsani, A. (2015b). Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infection Genetics and Evolution* **30**, 278-287.

Dekker, E. L., Woolston, C. J., Xue, Y. B., Cox, B. & Mullineaux, P. M. (1991). Transcript mapping reveals different expression strategies for the bicistronic RNAs of the geminivirus wheat dwarf virus. *Nucleic Acids Res* **19**, 4075-4081.

Delwart, E. & Li, L. (2012). Rapidly expanding genetic diversity and host range of the Circoviridae viral family and other Rep encoding small circular ssDNA genomes. *Virus Res* **164**, 114-121.

Diemer, G. S. & Stedman, K. M. (2012). A novel virus genome discovered in an extreme environment suggests recombination between unrelated groups of RNA and DNA viruses. *Biol Direct* **7**, e13.

Du, Z., Tang, Y., Zhang, S., She, X., Lan, G., Varsani, A. & He, Z. (2014). Identification and molecular characterization of a single-stranded circular DNA virus with similarities to Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1. *Archives of virology* **159**, 1527-1531.

Dunlap, D. S., Ng, T. F. F., Rosario, K., Barbosa, J. G., Greco, A. M., Breitbart, M. & Hewson, I. (2013). Molecular and microscopic evidence of viruses in marine copepods. *P Natl Acad Sci USA* **110**, 1375-1380.

Fahsbender, E., Hewson, I., Rosario, K., Tuttle, A. D., Varsani, A. & Breitbart, M. (2015). Discovery of a novel circular DNA virus in the Forbes sea star, Asterias forbesi. *Archives of virology* **160**, 2349-2351.

Garigliany, M.-M., Hagen, R. M., Frickmann, H., May, J., Schwarz, N. G., Perse, A., Jöst, H., Börstler, J., Shahhosseini, N. & Desmecht, D. (2014). Cyclovirus CyCV-VN species distribution is not limited to Vietnam and extends to Africa. *Scientific reports* **4**.

Garigliany, M. M., Boerstler, J., Joest, H., Badusche, M., Desmecht, D., Schmidt-Chanasit, J. & Cadar, D. (2015). Characterization of a novel circo-like virus in Aedes vexans mosquitoes from Germany: evidence for a new genus within the family Circoviridae. *J Gen Virol* **96**, 915-920.

Ge, X., Li, J., Peng, C., Wu, L., Yang, X., Wu, Y., Zhang, Y. & Shi, Z. (2011). Genetic diversity of novel circular ssDNA viruses in bats in China. *J Gen Virol* **92**, 2646-2653.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y. & Shi, Z. (2012). Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *Journal of Virology* **86**, 4620-4630.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321.

Han, H. J., Wen, H. L., Zhou, C. M., Chen, F. F., Luo, L. M., Liu, J. W. & Yu, X. J. (2015). Bats as reservoirs of severe emerging infectious diseases. *Virus Res* **205**, 1-6.

Hanna, Z. R., Runckel, C., Fuchs, J., DeRisi, J. L., Mindell, D. P., Van Hemert, C., Handel, C. M. & Dumbacher, J. P. (2015). Isolation of a complete circular virus genome sequence from an Alaskan black-capped chickadee (Poecile atricapillus) gastrointestinal tract sample. *Genome announcements* **3**, 01081-01015.

Hansen, T. A., Fridholm, H., Froslev, T. G., Kjartansdottir, K. R., Willerslev, E., Nielsen, L. P. & Hansen, A. J. (2015). New Type of Papillomavirus and Novel Circular Single Stranded DNA Virus Discovered in Urban Rattus norvegicus Using Circular DNA Enrichment and Metagenomics. *PLoS One* **10**, e0141952.

He, B., Li, Z., Yang, F., Zheng, J., Feng, Y., Guo, H., Li, Y., Wang, Y., Su, N., Zhang, F., Fan, Q. & Tu, C. (2013). Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel Mammalian viruses. *PLoS One* **8**, e61950.

Hewson, I., Eaglesham, J. B., Hook, T. O., LaBarre, B. A., Sepulveda, M. S., Thompson, P. D., Watkins, J. M. & Rudstam, L. G. (2013a). Investigation of viruses in Diporeia spp. from the Laurentian Great Lakes and Owasco Lake as potential stressors of declining populations. *J Great Lakes Res* **39**, 499-506.

Hewson, I., Ng, G., Li, W. F., LaBarre, B. A., Aguirre, I., Barbosa, J. G., Breitbart, M., Greco, A. W., Kearns, C. M., Looi, A., Schaffner, L. R., Thompson, P. D. & Hairston, N. G. (2013b). Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a Daphnia-associated single-stranded DNA virus in two temperate lakes. *Limnol Oceanogr* **58**, 1605-1620.

Horser, C. L., Karan, M., Harding, R. M. & Dale, J. L. (2001). Additional Rep-encoding DNAs associated with banana bunchy top virus. *Archives of virology* **146**, 71-86.

Katul, L., Timchenko, T., Gronenborn, B. & Vetten, H. J. (1998). Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *J Gen Virol* **79**, 3101-3109.

Kim, H. K., Park, S. J., Nguyen, V. G., Song, D. S., Moon, H. J., Kang, B. K. & Park, B. K. (2012). Identification of a novel single-stranded, circular DNA virus from bovine stool. *J Gen Virol* **93**, 635-639.

Kim, K. H., Chang, H. W., Nam, Y. D., Roh, S. W., Kim, M. S., Sung, Y., Jeon, C. O., Oh, H. M. & Bae, J. W. (2008). Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl Environ Microb* **74**, 5975-5985.

King, A. M., Adams, M. J. & Lefkowitz, E. J. (2011). *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*: Elsevier.

Kraberger, S., Arguello-Astorga, G. R., Greenfield, L. G., Galilee, C., Law, D., Martin, D. P. & Varsani, A. (2015a). Characterisation of a diverse range of circular replication-associated protein encoding DNA viruses recovered from a sewage treatment oxidation pond. *Infection Genetics and Evolution* **31C**, 73-86.

Kraberger, S., Farkas, K., Bernardo, P., Booker, C., Arguello-Astorga, G. R., Mesleard, F., Martin, D. P., Roumagnac, P. & Varsani, A. (2015b). Identification of novel Bromus- and Trifolium-associated circular DNA viruses. *Archives of virology* **160**, 1303-1311.

Kraberger, S., Stainton, D., Dayaram, A., Zawar-Reza, P., Gomez, C., Harding, J. S. & Varsani, A. (2013). Discovery of Sclerotinia sclerotiorum Hypovirulence-Associated Virus-1 in Urban River Sediments of Heathcote and Styx Rivers in Christchurch City, New Zealand. *Genome announcements* **1**, 00559-00513.

Labonte, J. M. & Suttle, C. A. (2013). Previously unknown and highly divergent ssDNA viruses populate the oceans. *ISME J* **7**, 2169-2177.

Lamberto, I., Gunst, K., Muller, H., Zur Hausen, H. & de Villiers, E. M. (2014). Mycovirus-like DNA virus sequences from cattle serum and human brain and serum samples from multiple sclerosis patients. *Genome announcements* **2**, 00848-00814.

Li, L., Giannitti, F., Low, J., Keyes, C., Ullmann, L. S., Deng, X., Aleman, M., Pesavento, P. A., Pusterla, N. & Delwart, E. (2015a). Exploring the virome of diseased horses. *J Gen Virol* **96**, 2721-2733.

Li, L., Victoria, J. G., Wang, C., Jones, M., Fellers, G. M., Kunz, T. H. & Delwart, E. (2010a). Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *Journal of Virology* **84**, 6955-6965.

Li, L. L., Kapoor, A., Slikas, B., Bamidele, O. S., Wang, C. L., Shaukat, S., Masroor, M. A., Wilson, M. L., Ndjango, J. B. N., Peeters, M., Gross-Camp, N. D., Muller, M. N., Hahn, B. H., Wolfe, N. D., Triki, H., Bartkus, J., Zaidi, S. Z. & Delwart, E. (2010b). Multiple Diverse Circoviruses Infect Farm Animals and Are Commonly Found in Human and Chimpanzee Feces. *Journal of Virology* **84**, 1674-1682.

Li, L. L., Shan, T. L., Soji, O. B., Alam, M. M., Kunz, T. H., Zaidi, S. Z. & Delwart, E. (2011). Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *J Gen Virol* **92**, 768-772.

Li, W., Gu, Y., Shen, Q., Yang, S., Wang, X., Wan, Y. & Zhang, W. (2015b). A novel gemycircularvirus from experimental rats. *Virus Genes* **51**, 302-305.

Lima, F. E., Cibulski, S. P., Dos Santos, H. F., Teixeira, T. F., Varela, A. P., Roehe, P. M., Delwart, E. & Franco, A. C. (2015). Genomic characterization of novel circular ssDNA viruses from insectivorous bats in Southern Brazil. *PLoS One* **10**, e0118070.

Liu, H., Fu, Y., Li, B., Yu, X., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Yi, X. & Jiang, D. (2011). Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC evolutionary biology* **11**, 276.

Liu, L., Davies, J. W. & Stanley, J. (1998). Mutational analysis of bean yellow dwarf virus, a geminivirus of the genus Mastrevirus that is adapted to dicotyledonous plants. *Journal of General Virology* **79**, 2265-2274.

Lorincz, M., Csagola, A., Farkas, S. L., Szekely, C. & Tuboly, T. (2011). First detection and analysis of a fish circovirus. *J Gen Virol* **92**, 1817-1821.

Male, M. F., Kami, V., Kraberger, S. & Varsani, A. (2015). Genome Sequences of Poaceae-Associated Gemycircularviruses from the Pacific Ocean Island of Tonga. *Genome announcements* **3**, 01144-01115.

Mansoor, S., Khan, S. H., Bashir, A., Saeed, M., Zafar, Y., Malik, K. A., Briddon, R., Stanley, J. & Markham, P. G. (1999). Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* **259**, 190-199.

McDaniel, L. D., Rosario, K., Breitbart, M. & Paul, J. H. (2014). Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences. *Environmental Microbiology* **16**, 570-585.

Miller, C. A. & Wilson, D. E. (1997). Pteropus tonganus. *Mammalian Species*, 1-6.

Moratelli, R. & Calisher, C. H. (2015). Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Memórias do Instituto Oswaldo Cruz* **110**, 1-22.

Morozov, I., Sirinarumitr, T., Sorden, S. D., Halbur, P. G., Morgan, M. K., Yoon, K. J. & Paul, P. S. (1998). Detection of a novel strain of porcine circovirus in pigs with postweaning multisystemic wasting syndrome. *Journal of clinical microbiology* **36**, 2535-2541.

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**, e108277.

Nash, T. E., Dallas, M. B., Reyes, M. I., Buhrman, G. K., Ascencio-Ibanez, J. T. & Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *Journal of Virology* **85**, 1182-1192.

Nelson, S. L., Kunz, T. H. & Humphrey, S. R. (2005). Folivory in fruit bats: leaves provide a natural source of calicum. *Journal of Chemical Ecology* **31**, 1683-1691.

Ng, T. F., Chen, L. F., Zhou, Y., Shapiro, B., Stiller, M., Heintzman, P. D., Varsani, A., Kondov, N. O., Wong, W., Deng, X., Andrews, T. D., Moorman, B. J., Meulendyk, T., MacKay, G., Gilbertson, R. L. & Delwart, E. (2014). Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch. *P Natl Acad Sci USA* **111**, 16842-16847.

Ng, T. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B. S., Wommack, K. E. & Delwart, E. (2012). High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *Journal of Virology* **86**, 12161-12175.

Ng, T. F. F., Alavandi, S., Varsani, A., Burghart, S. & Breitbart, M. (2013). Metagenomic identification of a nodavirus and a circular ssDNA virus in semi-purified viral nucleic acids from the hepatopancreas of healthy Farfantepenaeus duorarum shrimp. *Dis Aquat Organ* **105**, 237-242.

Ng, T. F. F., Willner, D. L., Lim, Y. W., Schmieder, R., Chau, B., Nilsson, C., Anthony, S., Ruan, Y. J., Rohwer, F. & Breitbart, M. (2011). Broad Surveys of DNA Viral Diversity Obtained through Viral Metagenomics of Mosquitoes. *PloS one* **6**, e20579.

Omatsu, T., Watanabe, S., Akashi, H. & Yoshikawa, Y. (2007). Biological characters of bats in relation to natural reservoir of emerging viruses. *Comp Immunol Microbiol Infect Dis* **30**, 357-374.

Padilla-Rodriguez, M., Rosario, K. & Breitbart, M. (2013). Novel cyclovirus discovered in the Florida woods cockroach *Eurycotis floridana* (Walker). *Archives of virology* **158**, 1389-1392.

Pham, H. T., Bergoin, M. & Tijssen, P. (2013). Acheta domesticus Volvovirus, a Novel Single-Stranded Circular DNA Virus of the House Cricket. *Genome announcements* **1**, 00079-00013.

Phan, T. G., Luchsinger, V., Avendano, L. F., Deng, X. & Delwart, E. (2014). Cyclovirus in nasopharyngeal aspirates of Chilean children with respiratory infections. *J Gen Virol* **95**, 922-927.

Phan, T. G., Mori, D., Deng, X., Rajindrajith, S., Ranawaka, U., Fan Ng, T. F., Bucardo-Rivera, F., Orlandi, P., Ahmed, K. & Delwart, E. (2015). Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**, 98-104.

Pie, J., Kim, B. & Grishin, N. (2008). PROMALS3D: a tool for multiple sequence and structure alignment. *Nucleic Acids Res* **36**, 2295-2300.

Pierson, E. D. & Rainey, W. E. (1992). The biology of flying foxes of the genus Pteropus: a review. In *Pacific island flying foxes: proceedings of an international conservation conference*, pp. 1-17: DTIC Document.

Plowright, R. K., Eby, P., Hudson, P. J., Smith, I. L., Westcott, D., Bryden, W. L., Middleton, D., Reid, P. A., McFarlane, R. A., Martin, G., Tabor, G. M., Skerratt, L. F., Anderson, D. L., Crameri, G., Quammen, D., Jordan, D., Freeman, P., Wang, L. F., Epstein, J. H., Marsh, G. A., Kung, N. Y. & McCallum, H. (2015). Ecological dynamics of emerging bat virus spillover. *Proc Biol Sci* **282**, 20142124.

Reavy, B., Swanson, M. M., Cock, P. J., Dawson, L., Freitag, T. E., Singh, B. K., Torrance, L., Mushegian, A. R. & Taliansky, M. (2015). Distinct circular single-stranded DNA viruses exist in different soil types. *Appl Environ Microb* **81**, 3934-3945.

Reuter, G., Boros, A., Delwart, E. & Pankovics, P. (2014). Novel circular single-stranded DNA virus from turkey faeces. *Archives of virology* **159**, 2161-2164.

Ritchie, B. W., Niagro, F. D., Lukert, P. D., Steffens, W. L., 3rd & Latimer, K. S. (1989). Characterization of a new virus from cockatoos with psittacine beak and feather disease. *Virology* **171**, 83-88.

Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M. & Varsani, A. (2012a). Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol* **93**, 2668-2681.

Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M. & Varsani, A. (2012b). Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol* **93**, 2668-2681.

Rosario, K., Duffy, S. & Breitbart, M. (2009). Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J Gen Virol* **90**, 2418-2424.

Rosario, K., Duffy, S. & Breitbart, M. (2012c). A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Archives of virology* **157**, 1851-1871.

Rosario, K., Marinov, M., Stainton, D., Kraberger, S., Wiltshire, E. J., Collings, D. A., Walters, M., Martin, D. P., Breitbart, M. & Varsani, A. (2011). Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). *J Gen Virol* **92**, 1302-1308.

Rosario, K., Schenck, R. O., Harbeitner, R. C., Lawler, S. N. & Breitbart, M. (2015). Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Front Microbiol* **6**, 0696.

Roux, S., Enault, F., Bronner, G., Vaulot, D., Forterre, P. & Krupovic, M. (2013). Chimeric viruses blur the borders between the major groups of eukaryotic single-stranded DNA viruses. *Nat Commun* **4**, e2700.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., Colombet, J., Sime-Ngando, T. & Debroas, D. (2012). Assessing the Diversity and Specificity of Two Freshwater Viral Communities through Metagenomics. *PloS one* **7**, e0033641.

Sachsenroder, J., Twardziok, S., Hammerl, J. A., Janczyk, P., Wrede, P., Hertwig, S. & Johne, R. (2012). Simultaneous Identification of DNA and RNA Viruses Present in Pig Faeces Using Process-Controlled Deep Sequencing. *PloS one* **7**, e0034631.

Sasaki, M., Orba, Y., Ueno, K., Ishii, A., Moonga, L., Hang'ombe, B. M., Mweene, A. S., Ito, K. & Sawa, H. (2015). Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses. *J Gen Virol* **96**, 440-452.

Sato, G., Kawashima, T., Kiuchi, M. & Tohya, Y. (2015). Novel cyclovirus detected in the intestinal contents of Taiwan squirrels (Callosciurus erythraeus thaiwanensis). *Virus Genes* **51**, 148-151.

Saunders, K. & Stanley, J. (1999). A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: Evidence for interfamilial recombination between plant DNA viruses. *Virology* **264**, 142-152.

Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A. & Delwart, E. (2011). The fecal virome of pigs on a high-density farm. *Journal of Virology* **85**, 11697-11708.

Sikorski, A., Dayaram, A. & Varsani, A. (2013a). Identification of a Novel Circular DNA Virus in New Zealand Fur Seal (Arctocephalus forsteri) Fecal Matter. *Genome announcements* **1**, 00558-00513.

Sikorski, A., Massaro, M., Kraberger, S., Young, L. M., Smalley, D., Martin, D. P. & Varsani, A. (2013b). Novel myco-like DNA viruses discovered in the faecal matter of various animals. *Virus Res* **177**, 209-216.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res* **19**, 1117-1123.

Smith, I. & Wang, L. F. (2013). Bats and their virome: an important source of emerging viruses capable of infecting humans. *Curr Opin Virol* **3**, 84-91.

Smith, R. J., Jeffries, T. C., Roudnew, B., Seymour, J. R., Fitch, A. J., Simons, K. L., Speck, P. G., Newton, K., Brown, M. H. & Mitchell, J. G. (2013). Confined aquifers as viral reservoirs. *Env Microbiol Rep* **5**, 725-730.

Smits, S. L., Zijlstra, E. E., van Hellemond, J. J., Schapendonk, C. M., Bodewes, R., Schurch, A. C., Haagmans, B. L. & Osterhaus, A. D. (2013). Novel cyclovirus in human cerebrospinal fluid, Malawi, 2010-2011. *Emerg Infect Dis* **19**, 1511.

Soffer, N., Brandt, M. E., Correa, A. M., Smith, T. B. & Thurber, R. V. (2014). Potential role of viruses in white plague coral disease. *The ISME journal* **8**, 271-283.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* **93**, 1103-1119.

Stainton, D., Martin, D. P., Collings, D. A., Thomas, J. E. & Varsani, A. (2016). Identification of defective molecules associated with isolates of Banana bunchy top virus. *Archives of virology*, DOI: 10.1007/s00705-00015-02736-00700.

Stainton, D., Martin, D. P., Muhire, B. M., Lolohea, S., Halafihi, M. i., Lepoint, P., Blomme, G., Crew, K. S., Sharman, M., Kraberger, S., Dayaram, A., Walters, M., Collings, D. A., Mabvakure, B., Lemey, P., Harkins, G. W., Thomas, J. E. & Varsani, A. (2015). The global distribution of Banana bunchy top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events. *Virus Evolution* **1**, vev009.

Tan, L. V., van Doorn, H. R., Nghia, H. D. T., Chau, T. T. H., Tu, L. T. P., de Vries, M., Canuti, M., Deijs, M., Jebbink, M. F., Baker, S., Bryant, J. E., Tham, N. T., Bkrong, N. T. T. C., Boni, M. F., Loi, T. Q., Phuong, L. T., Verhoeven, J. T. P., Crusat, M., Jeeninga, R. E., Schultsz, C., Chau, N. V. V., Hien, T. T., van der Hoek, L., Farrar, J. & de Jong, M. D. (2013). Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *mBio* **4**, 00231-00213.

van den Brand, J. M., van Leeuwen, M., Schapendonk, C. M., Simon, J. H., Haagmans, B. L., Osterhaus, A. D. & Smits, S. L. (2012). Metagenomic analysis of the viral flora of pine marten and European badger feces. *Journal of Virology* **86**, 2360-2365.

Varsani, A., Kraberger, S., Jennings, S., Porzig, E. L., Julian, L., Massaro, M., Pollard, A., Ballard, G. & Ainley, D. G. (2014a). A novel papillomavirus in Adelie penguin (Pygoscelis adeliae) faeces sampled at the Cape Crozier colony, Antarctica. *J Gen Virol* **95**, 1352-1365.

Varsani, A., Navas-Castillo, J., Moriones, E., Hernandez-Zepeda, C., Idris, A., Brown, J. K., Murilo Zerbini, F. & Martin, D. P. (2014b). Establishment of three new

genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch Virol* **159**, 2193-2203.

Varsani, A., Porzig, E. L., Jennings, S., Kraberger, S., Farkas, K., Julian, L., Masaro, M., Ballard, G. & Ainley, D. G. (2015). Identification of an avian'polyomavirus associated with Adelie penguins (Pygoscelis adeliae). *J Gen Virol* **96**, 851-857.

Whon, T. W., Kim, M. S., Roh, S. W., Shin, N. R., Lee, H. W. & Bae, J. W. (2012). Metagenomic Characterization of Airborne Viral DNA Diversity in the Near-Surface Atmosphere. *Journal of Virology* **86**, 8221-8231.

Wong, S., Lau, S., Woo, P. & Yuen, K. Y. (2007). Bats as a continuing source of emerging infections in humans. *Rev Med Virol* **17**, 67-91.

Woo, P. C., Lau, S. K., Teng, J. L., Tsang, A. K., Joseph, M., Wong, E. Y., Tang, Y., Sivakumar, S., Bai, R., Wernery, R., Wernery, U. & Yuen, K. Y. (2014). Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses. *Virology* **471-473**, 117-125.

Wright, E., Heckel, T., Groenendijk, J., Davies, J. & Boulton, M. (1997). Splicing features in maize streak virus virion-and complementary-sense gene expression. *The plant journal* **12**, 1285-1297.

Wu, R. Y., You, L. R. & Soong, T. S. (1994). Nucleotide sequences of two circular single-stranded DNAs associated with Banana bunchy top virus. *Phytopathology* **84**, 952-958.

Wu, Z., Ren, X., Yang, L., Hu, Y., Yang, J., He, G., Zhang, J., Dong, J., Sun, L., Du, J., Liu, L., Xue, Y., Wang, J., Yang, F., Zhang, S. & Jin, Q. (2012). Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *Journal of Virology* **86**, 10999-11012.

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S. & Jin, Q. (2015). Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *The ISME journal*, doi: 10.1038/ismej.2015.1138.

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T. & Takai, K. (2013). Metagenomic analysis of viral communities in (hado)pelagic sediments. *PLoS One* **8**, e57271.

Yu, X., Li, B., Fu, Y. P., Jiang, D. H., Ghabrial, S. A., Li, G. Q., Peng, Y. L., Xie, J. T., Cheng, J. S., Huang, J. B. & Yi, X. H. (2010). A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *P Natl Acad Sci USA* **107**, 8387-8392.

Zawar-Reza, P., Arguello-Astorga, G. R., Kraberger, S., Julian, L., Stainton, D., Broady, P. A. & Varsani, A. (2014). Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infection Genetics and Evolution* **26**, 132-138.

Zhang, W., Li, L., Deng, X., Kapusinszky, B., Pesavento, P. A. & Delwart, E. (2014). Faecal virome of cats in an animal shelter. *J Gen Virol* **95**, 2553-2564.

Zhou, X. P. (2013). Advances in Understanding Begomovirus Satellites. *Annu Rev Phytopathol* **51**, 357-381.

Chapter 3

Conclusion

Contents

3.1 Rationale behind this study

As part of a continuing effort to explore the diversity of CRESS DNA viruses circulating in the Pacific region, this viral metagenomics thesis research project was undertaken to identify CRESS DNA viruses from faeces of Pacific flying foxes (*P. tonganus*). The faecal samples were collected from four roosting sites (Ha'ateiho ('Atele), Lapaha (Takuilau), Ha'avakatolo and Kolovai) located in Tongatapu the main island of Tonga in 2014 and 2015. Various studies have demonstrated that viral metagenomics studies have shown a high abundance of CRESS DNA viruses in animal faecal matter. Animal faeces sampling is non-invasive and it may contain a wide range of viruses including those associated with the food they eat and the environment of the animal.

The objective of this thesis research was to 1) identify CRESS DNA viruses associated with *P. tonganus* faecal matter; 2) determine whether there are differences in CRESS DNA viral assemblages at four *P. tonganus* roosting sites on Tongatapu; 3) determine whether any CRESS DNA viruses are persistently associated with *P. tonganus* faeces; 4) identify putative plant-infecting viruses circulating in Tonga via faecal sampling given that *P. tonganus* are frugivores.

3.2 Diversity of CRESS DNA viruses in Tonga

Very little is known about CRESS DNA viruses circulating in the Tongan archipelago and to date only a handful of these have been identified. Prior to this study, only six species of CRESS-DNA viruses had been previously identified. These include a nanovirus (*Banana bunchy top virus*) (Stainton *et al.*, 2015; Stainton *et al.*, 2012) infecting bananas, a cyclovirus (Dragonfly cyclovirus - 1) recovered from *Diplacodes bipunctata*, *Pantala flavescens*, *Tholymis tillarga* (Rosario & Breitbart, 2011; Rosario *et al.*, 2012), two gemycircularviruses (Poaceae associated gemycircularvirus - 1 recovered from *Brachiaria deflexa*, *Saccharum hybrid* and Dragonfly associated circular virus 3 from *P. flavescens*) (Male *et al.*, 2015; Rosario *et al.*, 2012) and two unclassified CRESS DNA viruses (Dragonfly circularisvirus from *P. flavescens* and Dragonfly orbiculatusvirus from *D. bipunctata*) (Rosario *et al.*, 2012) (Table 3.1).

Table 3.1: A summary of all the CRESS DNA viruses that were previously identified in Tonga together with those recovered from this study.

CRESS DNA virus type	Description	Acronyms	Collection year	Host	Sample type	References
Nanoviruses	*Banana bunchy top virus* (n=77)	BBTV	1989-2010	*Musa* spp	Banana leaf	Stainton et al., 2012, 2015 & Karan et al., 1994
Cycloviruses	Dragonfly cyclovirus 1 (n=24)	DfCyV-1	2010	*Tholymis tillarga, Diplacodes bipunctat, Pantala flavescens*	Dragonfly abdomen	Rosario et al., 2011, 2012a
	Pacific flying fox faeces associated cyclovirus-1	**PfffaCyV-1**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated cyclovirus-2	**PfffaCyV-2**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated cyclovirus-3 (n=3)	**PfffaCyV-3**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
Gemycircularviruses	Dragonfly-associated circular virus 3	DfasCV-3	2010	*Pantala flavescens*	Dragonfly abdomen	Rosario et al., 2012a
	Poaceae associated gemycircularvirus-1 (n=3)	PaGmV-1	2014	*Brachiaria deflexa, Saccharum hybrid*	Signalgrass	Male et al., 2015
	Pacific flying fox faeces associated gemycircularvirus-1 (n=2)	**PfffaGmV-1**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-2 (n=2)	**PfffaGmV-2**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-3	**PfffaGmV-3**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-4 (n=2)	**PfffaGmV-4**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-5	**PfffaGmV-5**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-6 (n=2)	**PfffaGmV-6**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-7	**PfffaGmV-7**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-8 (n=2)	**PfffaGmV-8**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-9	**PfffaGmV-9**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-10 (n=2)	**PfffaGmV-10**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-11 (n=6)	**PfffaGmV-11**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-12	**PfffaGmV-12**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-13	**PfffaGmV-13**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying fox faeces associated gemycircularvirus-14	**PfffaGmV-14**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
Unclassified CRESS DNA viruses	Dragonfly circularisvirus	DfCirV	2010	*Pantala flavescens*	Dragonfly abdomen	Rosario et al., 2012a
	Dragonfly orbiculatusvirus	DfOrV	2010	*Diplacodes bipunctata*	Dragonfly abdomen	Rosario et al., 2012a
	Pacific flying foxes faeces asociated circular virus-1 (n=2)	**PfffaCV-1**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-2 (n=2)	**PfffaCV-2**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-3	**PfffaCV-3**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-4	**PfffaCV-4**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-5	**PfffaCV-5**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-6	**PfffaCV-6**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-7	**PfffaCV-7**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-8	**PfffaCV-8**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-9	**PfffaCV-9**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-10	**PfffaCV-10**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-11	**PfffaCV-11**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-12	**PfffaCV-12**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-13	**PfffaCV-12**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-14	**PfffaCV-12**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
	Pacific flying foxes faeces asociated circular virus-15	**PfffaCV-12**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**
Multi-component virus	**Pacific flying foxes faeces asociated multicomponent virus-1**	**PfffaMV-1**	**2014-2015**	***Pteropus tonganus***	**Bat faeces**	**This study**

3.3 Summary of the results

The viral metagenomics study of Pacific flying fox faeces, where the faeces were sampled from four roosting sites in Tongatapu, identified five novel cycloviruses representing three putative species, 25 gemycircularviruses representing at least 14 putative species, 17 unclassified CRESS DNA viruses representing 15 putative species and a putative novel multi-component virus with three cognate molecules (Table 3.1). Two circular DNA molecules were also identified in this study which may represent either defective genomes or individual components of multi-component genomes of CRESS DNA viruses. The data obtained from this study has significantly expanded the diversity of CRESS DNA viruses that are circulating in Tonga.

This study shows that there are differences in CRESS DNA viral assemblages at the four *P. tonganus* roosting sites in Tongatapu. In total, 20 CRESS DNA viral species were recovered from the Ha'avakatolo Pacific flying fox roosting site, 13 from Ha'ateiho ('Atele) and two each from Kolovai and Lapaha (Takuilau). Given that Kolovai and Ha'avakatolo are neighbouring villages, it is surprising that the *P. tonganus* roosting colonies of these two villages only had one viral species in common. The other viral species identified in Kolovai (PfffaMCV-1) was not found in Ha'avakatolo or in fact in any of the other sites. However, although Ha'avakatolo and Ha'ateiho are geographically more distant, they share six of the same virus species.

Several CRESS DNA viral species were identified in both 2014 and 2015 suggesting that these viruses are persistently associated with faecal matter of Pacific flying foxes. A number of viruses were also identified in more than one sampling site in 2015 suggesting these viruses have a broad distribution across the island amongst the Pacific flying fox colonies. These viruses are possibly common in Pacific flying fox roosting colonies but whether the Pacific flying foxes in Tonga move between roosting sites or roost at specific colonies through their lifetime remains unknown.

PfffaCyV-3s share high similarity (~77% genome-wide identity) with the Human cyclovirus VS5700009 (KC771281) which was identified in a patient with paraplegia of unknown aetiology in Malawi (Smits *et al.*, 2013). Taking into consideration the observations by Garigliany *et al.* (2014) and Cotmore *et al.* (2014) where they found cyclovirus CyCN-VN in different geographical locations (Africa and Asia) and in humans and domestic animals, it is

possible that the cycloviruses identified in Pacific flying foxes, and more so PfffaCyV-3 which is most closely related to CyCN-VN, may be found associated with other animals in Tonga.

Given that *P. tonganus* are frugivores, it was expected that a number of plant-infecting viruses that are associated with plants would be identified via faecal sampling. None of the viruses identified in this study are known plant-infecting viruses. Interestingly, a multi-component virus (PfffaMV-1) recovered from this study share no similarities with the members of the *Nanoviridae* family, or with alphasatellite molecules associated with either members of *Nanoviridae* or *Geminiviridae* families. Currently all classified ssDNA multi-component viruses are known to infect plants therefore PfffaMV-1 could be a putative plant-infecting virus.

3.4 Significance of this study and future directions

Viral metagenomics can assist with a better understanding of viral ecology biogeography, revealing viral diversity and can be used for virus surveillance of emerging pathogens, as reviewed in Rosario and Breitbart (2011). Therefore, within a Tongan context perhaps identifying zoonotic viruses may be important and hence a viral metagenomics study of RNA and DNA viruses in the Pacific flying fox faeces may potentially be carried out in the future. Furthermore, given that insect vectors such as mosquitoes play such a crucial role in viral transmission to humans and animals, a metagenomic study of viruses in these vectors would be of public health importance.

This study has paved the way for more targeted studies to be undertaken in Tonga. The viral sequences from this study can be used to design specific probes to identify whether these viruses are circulating in humans, wild animals and possibly plants. Overall, this study has identified unique viral sequences in the Tongan archipelago and provided a small snapshot of the CRESS DNA viral diversity associated with Pacific flying fox faeces. From an academic point of view, it would be ideal to undertake a viral metagenomics study of Pacific flying fox faeces throughout the Pacific in an attempt to 1) identify CRESS DNA viruses associated with their faeces; 2) identify whether any of these viruses are geographically limited; 3) identify viruses that are frequently associated with these bats across the Pacific; 4) carry out infectivity studies of these viruses. Without a doubt, it is evident that viral metagenomics studies of bat faecal matter will enable a rapid exploration of viral diversity.

3.5 References

Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. A., Mukha, D. V., Pintel, D. J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D. & Davison, A. J. (2014). The family Parvoviridae. *Arch Virol* **159**, 1239-1247.

Garigliany, M.-M., Hagen, R. M., Frickmann, H., May, J., Schwarz, N. G., Perse, A., Jöst, H., Börstler, J., Shahhosseini, N. & Desmecht, D. (2014). Cyclovirus CyCV-VN species distribution is not limited to Vietnam and extends to Africa. *Scientific reports* **4**.

Male, M. F., Kami, V., Kraberger, S. & Varsani, A. (2015). Genome Sequences of Poaceae-Associated Gemycircularviruses from the Pacific Ocean Island of Tonga. *Genome Announc* **3**, e01144-01115.

Rosario, K. & Breitbart, M. (2011). Exploring the viral world through metagenomics. *Curr Opin Virol* **1**, 289-297.

Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M. & Varsani, A. (2012). Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J Gen Virol* **93**, 2668-2681.

Smits, S. L., Zijlstra, E. E., van Hellemond, J. J., Schapendonk, C. M., Bodewes, R., Schurch, A. C., Haagmans, B. L. & Osterhaus, A. D. (2013). Novel cyclovirus in human cerebrospinal fluid, Malawi, 2010-2011. *Emerg Infect Dis* **19**, 1511.

Stainton, D., Halafihi, M., Collings, D. A. & Varsani, A. (2015). Genome Sequence of Banana Streak MY Virus from the Pacific Ocean Island of Tonga. *Genome Announc* **3**, e00543-00515.

Stainton, D., Kraberger, S., Walters, M., Wiltshire, E. J., Rosario, K., Halafihi, M., Lolohea, S., Katoa, I., Faitua, T. H., Aholelei, W., Taufa, L., Thomas, J. E., Collings, D. A., Martin, D. P. & Varsani, A. (2012). Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *J Gen Virol* **93**, 1103-1119.