Enhancing Arbovirus Surveillance with Metatranscriptomics

Submitted by

Jana Batovska

Bachelor of Biomedical Science (Honours)

Submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Applied Systems Biology

College of Science, Health and Engineering

La Trobe University

Victoria, Australia

June 2020

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

%CC / %CC-r	Percent coverage by contigs / percent coverage by contigs ratio
%CR / %CR-r	Percent coverage by reads / percent coverage by reads ratio
+ssRNA / -ssRNA	Positive-sense / negative-sense single-stranded RNA
AA PI%	Amino acid percentage identity
AgDNV	Anopheles gambiae densovirus
AKAV	Akabane virus
AMIV	Anopheles minimus iridovirus
APRV	Aedes pseudoscutellaris reovirus
AUD	Australian Dollar
BAV	Banna virus
BEFV	Bovine ephemeral fever virus
BFV	Barmah Forest virus
BGS	Biogents Sentinel
BLAST	Basic Local Alignment Search Tool
BOLD	Barcode of Life Data System
bp	Base pair
BTV	Bluetongue virus
CCHFV	Crimean-Congo hemorrhagic fever virus
cDNA	Complementary DNA
CDNA	Communicable Diseases Network Australia
CHIKV	Chikungunya virus
COI	Cytochrome oxidase I
Contig	Contiguous sequence
CORV	Corriparta virus
Ct	Cycle threshold
CuniNPV	Culex nigripalpus nucleopolyhedrovirus

CxFV	Culex flavivirus
ddPCR	Droplet digital polymerase chain reaction
DENV	Dengue virus
dsDNA / dsRNA	Double-stranded DNA / double-stranded RNA
EILV	Eilat alphavirus
ELISA	Enzyme-linked immunosorbent assay
EVS	Encephalitis virus surveillance
FCR / FCR-r	Fold coverage by reads / fold coverage by reads ratio
FSV1	Fisavirus 1
FTA	Flinders Technology Associates
GTR	General time reversible
HBV	Hepatitis B virus
HMDV	Hypsignathus monstrosus dicistrovirus
HTS	High-throughput sequencing
IFA	Immunofluorescence assay
InDA-C	Insert dependent adaptor cleavage
ICTV	The International Committee on Taxonomy of Viruses
ISV	Insect-specific virus
JEV	Japanese encephalitis virus
kbp	Kilobase pair
KHV	Koyama Hill virus
KUNV	West Nile virus Kunjin subtype
LLSV	Llano Seco virus
mAbs	Monoclonal antibodies
MAVRIC	Monoclonal antibodies against viral RNA intermediates in cell culture
MINV	Minnal virus
mRNA	Messenger RNA

MSeq	Metatranscriptomic sequencing
MSV	Mosquito-specific virus
MVEV	Murray Valley encephalitis virus
NAMAC	National Arbovirus and Malaria Advisory Committee
NCBI	National Center for Biotechnology Information
NETV	Netivot virus
NGS	Next-generation sequencing
NMDS	Non-metric multidimensional scaling
NNDSS	National Notifiable Diseases Surveillance System
nr	Non-redundant
NRIV	Ngari virus
NSW	New South Wales
nt	Nucleotide
NTC	No-template control
OBOV	Obodhiang ephemerovirus
ONNV	O'nyong-nyong virus
ONT	Oxford Nanopore Technologies
ORF	Open reading frame
PacBio	Pacific Biosciences
PBT	Passive box traps
PCC / PCC-r	Percent coverage by contigs / percent coverage by contigs ratio
PIRYV	Piry virus
QLD	Queensland
qPCR	Quantitative polymerase chain reaction
RdRp	RNA-dependent RNA polymerase
RPM / RPM-r	Reads per million / reads per million ratio
rRNA	Ribosomal RNA

RRV	Ross River virus
RT-ddPCR	Reverse transcription droplet digital polymerase chain reaction
RT-LAMP	Reverse transcription loop-mediated isothermal amplification
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
RVFV	Rift Valley fever virus
SINV	Sindbis virus
SLEV	St. Louis encephalitis virus
SLOV	Stretch Lagoon orbivirus
SMACK	Sentinel mosquito arbovirus capture kits
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
ssDNA	Single-stranded DNA
TRUV	Trubanaman virus
UMAV	Umatilla virus
USA	United States of America
UsCPV	Uranotaenia sapphirina cypovirus
USD	United States Dollar
UTR	Untranslated region
VADCP	Victorian Arbovirus Disease Control Program
WA	Western Australia
WEEV	Western equine encephalitis virus
WGRV	Wongorr virus
WNV	West Nile virus
WTGs	Whole trap grinds
YFV	Yellow fever virus
YYV	Yada Yada virus
ZIKV	Zika virus

SUMMARY

Mosquitoes contribute to the global transmission of arboviruses that cause serious diseases in humans and animals. Surveillance programs help predict and control the spread of these diseases by monitoring arbovirus activity in mosquito populations. Viral metatranscriptomic sequencing can enhance surveillance by providing untargeted, high-throughput arbovirus detection with whole genome information from unsorted, field-collected mosquitoes. In order to enable the use of metatranscriptomic sequencing in routine arbovirus surveillance activities, this thesis details the development and optimisation of a mosquito-specific protocol for metatranscriptomic arbovirus detection. Analytical sensitivity and specificity of the metatranscriptomic protocol was determined via a spiking experiment and a positive detection criterion was established based on sequence coverage of the arbovirus genome. The metatranscriptomic protocol was applied to unsorted, bulk mosquito traps collected longitudinally in various locations, leading to the detection of five arboviruses of relevance to public health. Comparison to established arbovirus detection methods indicated that, while PCR-based approaches were more sensitive, metatranscriptomics provided more genomic information, did not require *a priori* knowledge of viruses, and allowed whole trap screening. Additionally, the insect species composition of the traps could be elucidated from the metatranscriptomic data, providing valuable information on vector species prevalence. Further analysis of the sequencing data gave insight into the previously unexplored viral diversity of mosquitoes from south-east Australia and led to the discovery of the first mosquito-specific alphavirus in the Asia-Pacific region, named Yada Yada virus. Long-read sequencing was also investigated to expand surveillance capabilities, with mosquito-based metatranscriptomic arbovirus detection demonstrated for the first time on the MinION sequencer using a Ross River virus-infected mosquito. The investigation of metatranscriptomic sequencing for arbovirus detection detailed in this thesis has illustrated how mosquito-based arbovirus surveillance can be upscaled, enhanced and expanded, and serves as an important resource for the incorporation of metatranscriptomics into routine arbovirus surveillance programs.

STATEMENT OF AUTHORSHIP

This thesis consists primarily of work by the author that has been published or accepted for publication as described in the text. Except where reference is made in the text of the thesis, this thesis contains no other material published elsewhere or extracted in whole or in part from a thesis submitted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Jana Batovska

22 June 2020

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Stacey Lynch for giving me the opportunity to do this PhD and hiring me all those years ago. Your enthusiasm and support have been invaluable to me, and I am grateful to have you as a role model. I would also like to thank my other supervisors: Brendan Rodoni for your openness and passion (giddy up!), Tim Sawbridge for guiding me through the mysteries of RNA-seq and asking the right questions, and Peter Mee, my fellow bin chook, for helping me with so much – I honestly didn't think I was going to be able to finish this PhD until you came along. Thank you all for your enduring patience and understanding, it has been a pleasure to work with you.

I am very grateful to have had the continued guidance and support of Mark Blacket, who gave me an appreciation of entomology and helped keep my head above water on several occasions, and Noel Cogan, who provided me with tools, resources, ideas, and space, and imparted upon me the wisdom to sequence everything. Your advice and generosity have meant a lot to me.

Many thanks to Helen Anderson for her help with the ddPCR, Michelle Drayton for her help in the lab, and Tracie Webster for sequencing so many samples and helping with my demultiplexing and spiking requests. I would also like to thank the BASC Team, past and present, for all their bioinformatics assistance. And a massive thank you goes to the amazing Arbovirology Team for helping with so many aspects of my experiments.

Linda Zheng, Rachel Mann, Jacqui Morris, and Elisse Nogarotto – thank you for all of the advice, help, chats, snacks, and hangs, you're all absolute legends. I would also like to thank Alex Piper for graciously accepting my endless requests for help, and Helen Hayden for reminding me to RTFM. And of course, thank you to the Micro Morning Tea Gang for all the delicious food and chats.

Thank you to all of the collaborators that have contributed to this PhD, especially Eddie Holmes, Mang Shi, and Jan Buchmann – my PhD journey started and ended with visiting your lab, with a whole bunch of yada yada in between. You helped teach me the ins and outs of metatranscriptomics, and for that I am deeply appreciative.

I would also like to thank all of the labs that hosted me in the USA. Joe Russell, Mark Zeller, Eric Delwart, Hanna Retallack, Lark Coffey – you all made me feel welcome and I appreciate you sharing your knowledge and experience with me. Thank you to Agriculture Victoria Research and the American Society for Virology for providing travel grants to help fund my trip to the USA.

This work was supported by an Australian Government Research Training Program Scholarship. A complement to this scholarship was provided by Agriculture Victoria Research.

To the wonderful Janet Wheeler – thank you for getting me through the difficult times and always making me feel capable. Thank you also to Kendra Whiteman for her administrative support. To everyone else who has assisted me during my PhD – thank you.

Lastly, a very special thank you to all of my friends and family, and to Bromley Hornsby for sharing this entire journey with me and supporting me throughout – you're the best around.

DEDICATION

I dedicate this thesis to my cat Cash, who passed the month before my submission. He came into my life right before I started this journey and left right before I could finish. He was there with me through all the highs and lows, keeping me company throughout the long nights spent writing. He brought me an immense amount of love and joy, and I miss him dearly.

THESIS PREFACE

This thesis consists of six chapters, where the experimental research chapters have been written in the format of self-contained journal articles. The first chapter provides a review of the relevant literature and states the research aims. The next four chapters detail the experimental work, with Chapters 2, 4 and 5 presented as published articles, and Chapter 3 as a manuscript intended for journal submission. Each of these chapters includes a preface that describes how the research links to the other chapters, the publication details, and a statement of joint authorship confirming the contribution of the PhD candidate. Any supplementary material is included at the end of each chapter. The sixth chapter summarises the main findings, provides integrative conclusions, and identifies avenues for further research. The chapters that have been published employ the referencing style of the corresponding journal, whereas Chapters 1, 4 and 6 have consistent referencing. A separate bibliography is included at the end of each chapter. Due to the novelty of the findings, Chapters 4 and 5 were published in a different order than presented in the thesis, with the chronological order as follows: Chapter 5 → Chapter 2 → Chapter 4 → Chapter 3.

CHAPTER 1

Introduction

1.1 Mosquito arboviruses

Arthropod-borne viruses (arboviruses) are defined as viruses that can replicate in both blood-feeding arthropods and their vertebrate hosts, and are maintained in nature via biological transmission (Russell et al., 2013). Mosquitoes (Diptera: Culicidae) are common vectors of arboviruses and pose a significant threat to human and animal health by contributing to the spread of diseases such as dengue fever, chikungunya, Japanese encephalitis, yellow fever, equine encephalitis and in recent years, Zika. Mosquito-borne arboviruses represent a major global health burden, with dengue virus alone infecting an estimated 390 million people per year, resulting in an annual cost of USD $8.9 billion (Bhatt et al., 2013; Shepard et al., 2016). Arboviral infections often result in long-term physical and neurological impairment, and very few vaccines or effective treatments are available (LaBeaud et al., 2011). In Australia, diseases such as Ross River fever, Barmah Forest virus disease, and Murray Valley encephalitis are all caused by endemic arboviruses transmitted by mosquitoes (Russell and Dwyer, 2000). From July 2014 to June 2015 there were 12,849 human cases of mosquito-borne disease in Australia, with Barmah Forest virus (BFV) and Ross River virus (RRV) accounting for 83% of these cases (Knope et al., 2019).

1.1.1 Arboviruses

A large diversity of viruses has been detected in mosquitoes, with both RNA and DNA genomes (Table 1.1). RNA viruses make up the majority of arboviruses, and their variability is attributed to a high mutation rate caused by virus-encoded RNA polymerases lacking proofreading ability, resulting in error-prone replication (Elena and Sanjuán, 2005). The positive-sense, single-stranded RNA (+ssRNA) virus group contains many of the arboviruses that cause widespread human disease, including: dengue virus (DENV), chikungunya virus (CHIKV), Japanese encephalitis virus (JEV), yellow fever virus (YFV), West Nile virus (WNV) and Zika virus (ZIKV).

Table 1.1: Examples of viruses that have been detected in mosquitoes, categorised according to the Baltimore classification system for viruses. The * symbol indicates viruses that are, or have been, endemic to Australia, and the † symbol indicates which viruses are notifiable under the Australian National Notifiable Diseases Surveillance System (NNDSS). Viruses highlighted in blue have human or animal hosts, whereas viruses highlighted in green are insect-specific or have no identified vertebrate host. Italicised names indicate viral species, genera, and families officially recognised by the International Committee on Taxonomy of Viruses (ICTV).

Group	Family	Genus	Virus species/names	References
Group I (dsDNA)	*Iridoviridae*	*Chloriridovirus*	*Anopheles minimus iridovirus* (AMIV)	(Huang et al., 2015)
	Baculoviridae	*Deltabaculovirus*	*Culex nigripalpus nucleopolyhedrovirus* (CuniNPV)	(Andreadis et al., 2003)
Group II (ssDNA)	*Parvoviridae*	*Brevidensovirus*	Anopheles gambiae densovirus (AgDNV)	(Ren et al., 2008)
Group III (dsRNA)	*Reoviridae*	*Seadornavirus*	*Banna virus* (BAV)	(Liu et al., 2010)
		Cypovirus	Uranotaenia sapphirina cypovirus (UsCPV)	(Shapiro et al., 2005)
		Orbivirus	Stretch Lagoon orbivirus (SLOV)*	(Cowled et al., 2009)
			Corriparta virus (CORV)*	(Belaganahalli et al., 2013)
		Dinovernavirus	*Aedes pseudoscutellaris reovirus* (APRV)	(Attoui et al., 2005)
Group IV (+ssRNA)	*Flaviviridae*	*Flavivirus*	*West Nile virus* (WNV)†	(Campbell et al., 2002)
			West Nile virus Kunjin subtype (KUNV)* †	(Prow et al., 2016)
			Yellow fever virus (YFV)†	(Gardner and Ryman, 2010)
			Murray Valley encephalitis virus (MVEV)* †	(Selvey et al., 2014a)
			Japanese encephalitis virus (JEV)†	(van den Hurk et al., 2009)
			Dengue virus (DENV)* †	(Gulholm and Rawlinson, 2017)
			Culex flavivirus (CxFV)	(Hoshino et al., 2007)
			Zika virus (ZIKV)	(Gutiérrez-Bugallo et al., 2019)

	Togaviridae	*Alphavirus*	*Chikungunya virus* (CHIKV)†	(Lo Presti et al., 2016; Viennet et al., 2013)
			Sindbis virus (SINV)*	(Ling et al., 2019; Sammels et al., 1999)
			Ross River virus (RRV)* †	(Harley et al., 2001)
			Barmah Forest virus (BFV)* †	(Jacups et al., 2008a)
			O'nyong-nyong virus (ONNV)	(Rezza et al., 2017)
Group V (-ssRNA)	*Bunyaviridae*	*Phlebovirus*	*Rift Valley fever virus* (RVFV)	(Gaudreault et al., 2019)
		Orthobunyavirus	Ngari virus (NRIV)	(Dutuze et al., 2018)
			Trubanaman virus (TRUV)*	(Gauci et al., 2016)
		Nairovirus	*Crimean-Congo hemorrhagic fever virus* (CCHFV)	(Zivcec et al., 2016)
	Rhabdoviridae	*Vesiculovirus*	Piry virus (PIRYV)	(de Souza et al., 2016)
		Ephemerovirus	*Bovine ephemeral fever virus* (BEFV)*	(Walker, 2005)
			Obodhiang ephemerovirus (OBOV)	(Blasdell et al., 2012)

In addition to transmitting a variety of arboviruses to vertebrate hosts, mosquitoes harbour a diverse range of insect-specific viruses (ISVs). Although this group of viruses are referred to as "insect-specific", the majority have been discovered in mosquitoes and are sometimes referred to in the literature as "mosquito-specific viruses" (MSVs) (Öhlund et al., 2019). These viruses infect mosquitoes and can replicate in mosquito cell lines *in vitro*, but do not replicate in vertebrate cells. Evidence suggests ISVs are maintained in mosquito populations via species-specific vertical transmission (from parent to offspring) (Bolling et al., 2012; Lutomiah et al., 2007; Saiyasombat et al., 2011) and venereal transmission between males and females (Barreau et al., 1997). Horizontal transmission between larvae and pupae in aquatic environments may also play a role (Ajamma et al., 2018). The advent of sequence-based technologies has led to a steep increase in the number of ISVs discovered in recent years (Agboli et al., 2019; Calisher and Higgs, 2018). Due to their ancestral relationship with arboviruses, ISVs are valuable in understanding viral origins and investigating the evolution of host switching processes (Li et al., 2015; Öhlund et al., 2019). Furthermore, some ISVs affect vector competency, leading to reduced replication (Bolling et al., 2012; Hobson-Peters et al., 2013; Kenney et al., 2014) and transmission (Goenaga et al., 2015; Hall-Mendelin et al., 2016; Romo et al., 2018) of pathogenic arboviruses in mosquitoes. The ability of ISVs to modulate arbovirus transmission has prompted interest in their use as biocontrol agents (Hall and Hobson-Peters, 2018; Patterson et al., 2020), similar to *Wolbachia*, the bacterial endosymbiont that has been recently used to help control the transmission of DENV in mosquitoes (Hoffmann et al., 2011).

1.1.2 Vectorial capacity of mosquitoes

There are over 3,500 species of mosquito worldwide, however only a handful of these are arbovirus vectors (Besansky et al., 2003). In order to be transmitted by a mosquito, a virus must reach the mosquito midgut and replicate, be released into the hemocoel and travel to the salivary glands, and then replicate and reside in the saliva until the mosquito feeds again (Beerntsen et al., 2000). There are several anatomical barriers within a mosquito that can hinder this process, including the midgut infection barrier, midgut escape barrier, salivary gland infection barrier, and salivary gland escape barrier (Figure 1.1). Intrinsic

factors that can influence biological vector competency include mosquito genetics (Beerntsen et al., 2000; Ciota et al., 2018), antiviral defence mechanisms (Lee et al., 2019), and midgut microbiota (Dennison et al., 2014). There are also a variety of extrinsic factors that affect arboviral replication and transmission in mosquitoes such as environmental temperature, rainfall and humidity (Ciota et al., 2018; Waldock et al., 2013). As such, the detection of arboviruses in mosquitoes does not necessarily signify transmission (Table 1.2).

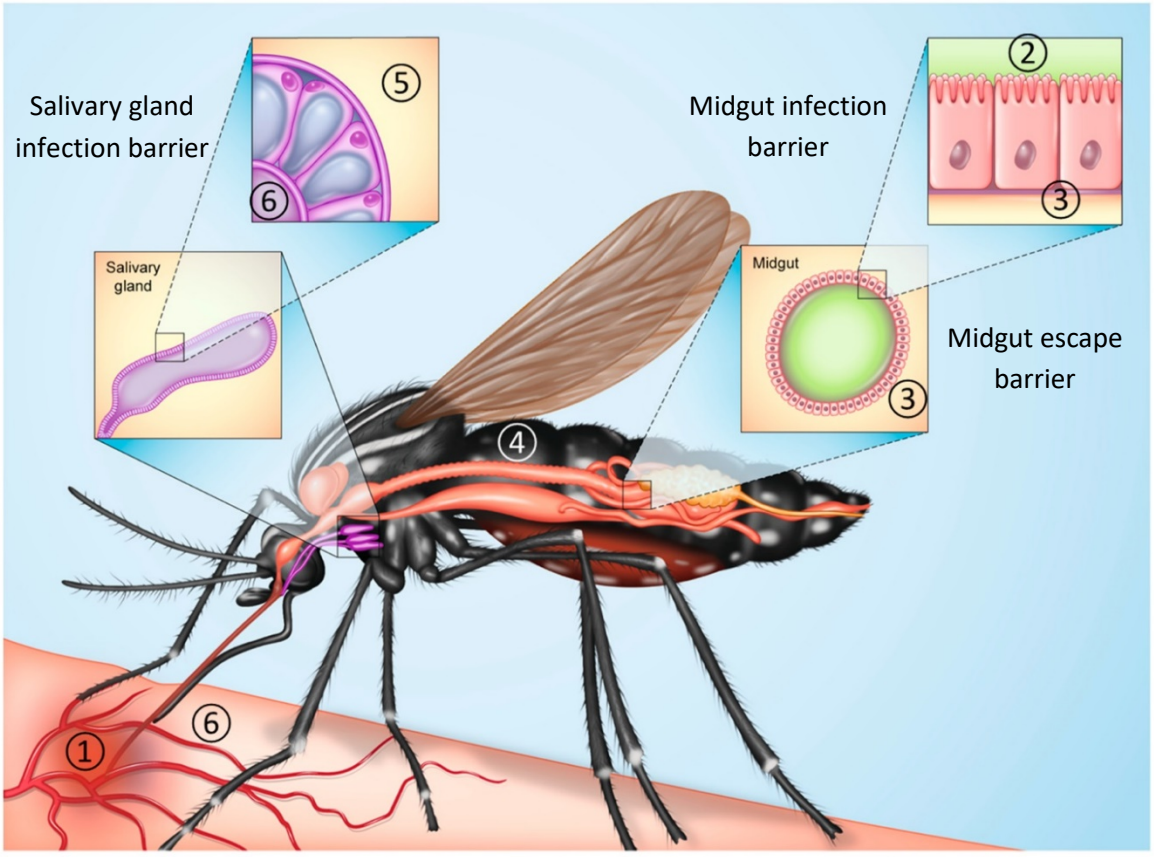


Figure 1.1: Different anatomical barriers to arbovirus transmission within a mosquito. **(1)** The mosquito feeds on an infected vertebrate host. **(2)** The virus travels to the midgut where it must infect and replicate in the midgut epithelial cells and **(3)** escape into the hemocoel. **(4)** Via the hemocoel the virus can infect the mosquito's peripheral tissues and organs. **(5)** In the salivary glands the virus must replicate and **(6)** infect the acinar cells, from which they then enter the saliva and are transmitted to a vertebrate host during the mosquito's next feeding. Adapted from Azar and Weaver (2019).

Besides being a biologically competent vector, mosquitoes need to feed on an infected host in order to acquire an arbovirus, which is then transmitted and maintained in complex cycles (Figure 1.2). Wild animal reservoirs serve as the primary hosts of arboviruses and may be re-infected several times, usually without exhibiting disease symptoms (Go et al., 2014). The natural transmission cycle between mosquitoes and wild animal reservoirs is referred to as the enzootic or sylvatic cycle, whereas transmission between mosquitoes and domestic animal hosts is the epizootic or rural cycle, and transmission between mosquitoes and humans is the epidemic or urban cycle (Weaver and Barrett, 2004). Zoonotic arboviruses can infect both animals and humans, with mosquitoes acting as 'bridge vectors' when they feed on animal reservoir hosts and then humans, resulting in spillover from enzootic and epizootic cycles (Armstrong and Andreadis, 2010). The urban or epidemic cycle occurs when humans develop viremia capable of re-infecting mosquitoes, which further propagates arbovirus transmission; however, if they do not contribute to the spread of an arbovirus, they are referred to as dead-end or incidental hosts (Figure 1.2). In addition to blood-feeding, mosquitoes can acquire viruses via vertical transmission (from adult to offspring) and venereal transmission (during mating). For instance, *Aedes aegypti* mosquitoes have been shown to vertically transmit DENV (Khin and Than, 1983), CHIKV (Agarwal et al., 2014), and ZIKV (Thangamani et al., 2016). The presence of arboviruses in mosquito eggs as a result of vertical transmission offers opportunities for egg and larvae-based surveillance (da Costa et al., 2017), however, results from field surveys suggest vertical transmission occurs at a low rate in nature and requires large sampling efforts to detect (Grunnill and Boots, 2016). Lastly, there is mechanical transmission, which does not require the virus to replicate inside the mosquito (Blanc and Gutiérrez, 2015).

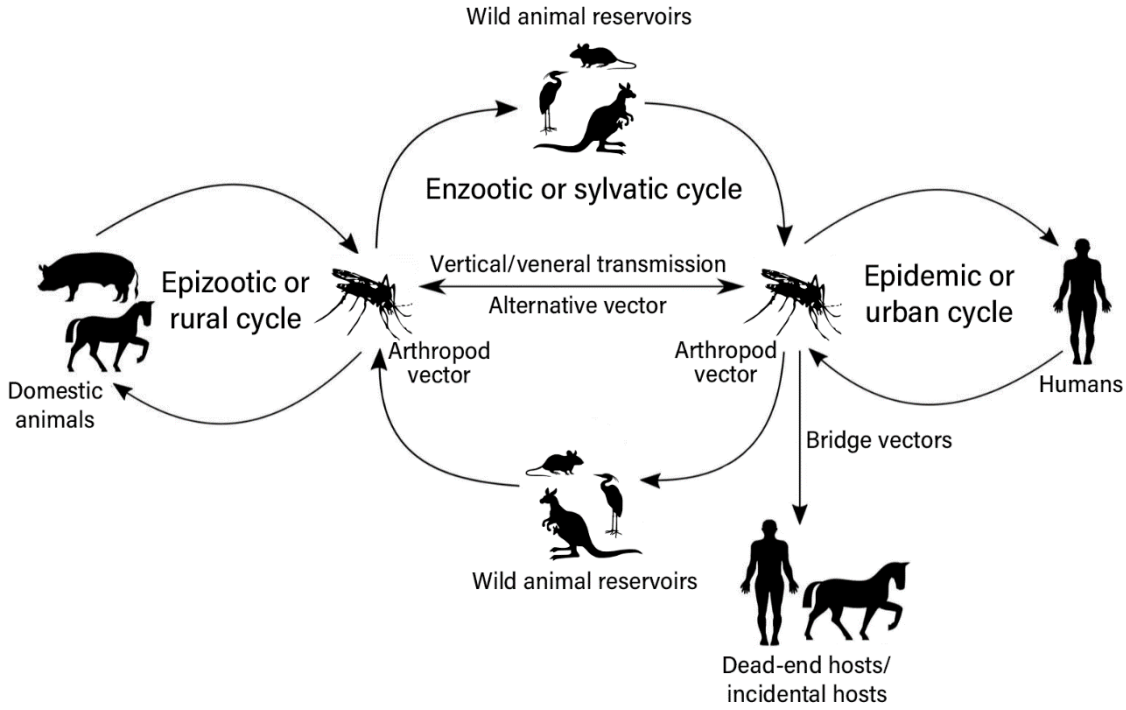


Figure 1.2: The different arbovirus transmission cycles. Adapted from Cholleti et al. (2018).

The feeding patterns of mosquitoes are species-specific, with some species feeding on a variety of hosts (generalist feeders), and others having a limited host range (specialist feeders) (Stephenson et al., 2019). Generalist feeders play a role in enzootic amplification of arboviruses due to their ability to acquire pathogens from animals and transmit them to humans, thereby acting as bridge vectors, whereas specialist feeders tend to contribute to only one cycle, such as the epidemic or urban transmission cycle or specific enzootic cycles (Figure 1.2).

Australia has over 300 mosquito species and only a minority of these contribute to arbovirus transmission (Webb et al., 2016). Some mosquito species have been shown to actively transmit several arboviruses, while others have only had arboviruses isolated from them (Table 1.2; Harley et al., 2001; Russell, 1996). *Aedes vigilax*, *Aedes notoscriptus* and *Culex annulirostris* all have a demonstrated capacity to transmit several arboviruses and are classified as generalist feeders (Stephenson et al., 2019), implicating them as primary arbovirus vector species in Australia.

Table 1.2: Arbovirus vector status for Australian mosquito species. Arboviruses marked with a 'D' have been detected in field-caught mosquitoes, whereas those marked with a 'T' have been shown to be transmitted by mosquitoes under experimental conditions. The mosquito species were chosen based on those listed in Gyawali et al. (2017).

Mosquito species	BFV	DENV	CHIKV	JEV	KUNV	MVEV	RRV	YFV [1]	References
Aedes aegypti		T	T	T	T	T	T	T	(van den Hurk et al., 2003, 2010, 2011; Kay, 1979; Watson and Kay, 1999)
Aedes alternans							D		(Lindsay et al., 1993; Russell et al., 1991)
Aedes bancroftianus	D						D		(Broom et al., 1993; Russell, 1986)
Aedes camptorhynchus	D						T		(Aldred et al., 1990; Ballard and Marshall, 1986)
Aedes clelandi							D		(Lindsay et al., 1992)
Aedes eidsvoldensis	D					D			(Broom et al., 1993; Mackenzie et al., 1994)
Aedes flavifrons							D		(McManus and Marshall, 1986)
Aedes normanensis	D					D	D		(Broom et al., 1989a, 1993; Doherty et al., 1979; Kay, 1979)
Aedes notoscriptus	T	T	T	T			T	T	(van den Hurk et al., 2003, 2010, 2011; Watson and Kay, 1998, 1999)
Aedes procax	T		T				T		(van den Hurk et al., 2010; Ryan and Kay, 1999; Ryan et al., 2000)
Aedes pseudonormanensis	D					D			(Broom et al., 1989b, 1993)
Aedes sagax						T	D		(Kay et al., 1989; Lindsay et al., 1992)
Aedes theobaldi							D		(Marshall, 1985)
Aedes tremulus					D	D	D		(Broom et al., 1995; Liehne et al., 1976; Lindsay et al., 1992)
Aedes vigilax	T		T	T		T	T		(Boyd and Kay, 1999; van den Hurk et al., 2003, 2010; Kay et al., 1979; Ryan et al., 2000)
Anopheles amictus	D						D		(Broom et al., 1993; Kay, 1979)
Anopheles annulipes	D					D	D		(Broom et al., 1993; Lindsay et al., 1992)
Anopheles bancroftii					D	D			(Doherty et al., 1968; Mackenzie et al., 1994)
Coquillettidia linealis	D		T				D		(van den Hurk et al., 2010; Lindsay et al., 1992; Mackenzie et al., 1994)

Culex annulirostris	T		T	T	T	T	T	(Boyd and Kay, 2000; van den Hurk et al., 2003, 2010; Kay et al., 1984, 1989; Ryan et al., 2000)
Culex australicus					D	D	T	(Marshall, 1979; Marshall et al., 1982; Ryan et al., 2000)
Culex palpalis						D		(Mackenzie et al., 1994)
Culex quinquefasciatus				T	D	T	D	(Doherty et al., 1979; van den Hurk et al., 2003; Lindsay et al., 1993; McLean, 1953)
Culex sitiens			T	T			D	(van den Hurk et al., 2003, 2010; Lindsay et al., 1993)
Mansonia uniformis				T		D	T	(van den Hurk et al., 2003; Mackenzie et al., 1994; Ryan et al., 2000)

[1] Abbreviations: BFV = Barmah Forest virus; DENV = dengue virus; CHIKV = chikungunya virus; JEV = Japanese encephalitis virus; KUNV = West Nile virus Kunjin subtype; MVEV = Murray Valley encephalitis virus; RRV = Ross River virus; YFV = yellow fever virus

1.1.3 Influence of climate on arbovirus transmission

Climatic factors such as rainfall, temperature, tides, and humidity can all impact mosquito abundance and distribution, and therefore influence arboviral activity (Russell, 1998). For instance, in Australia there is a range of climate zones that affect the transmission of arboviruses, with the hot and humid summers in the most northerly parts of the country conducive to outbreaks (Figure 1.3). The intrinsic relationship between mosquitoes, arboviruses and climate means that weather forecasting can be used for predictive outbreak modelling by arbovirus surveillance programs (Knope et al., 2019).

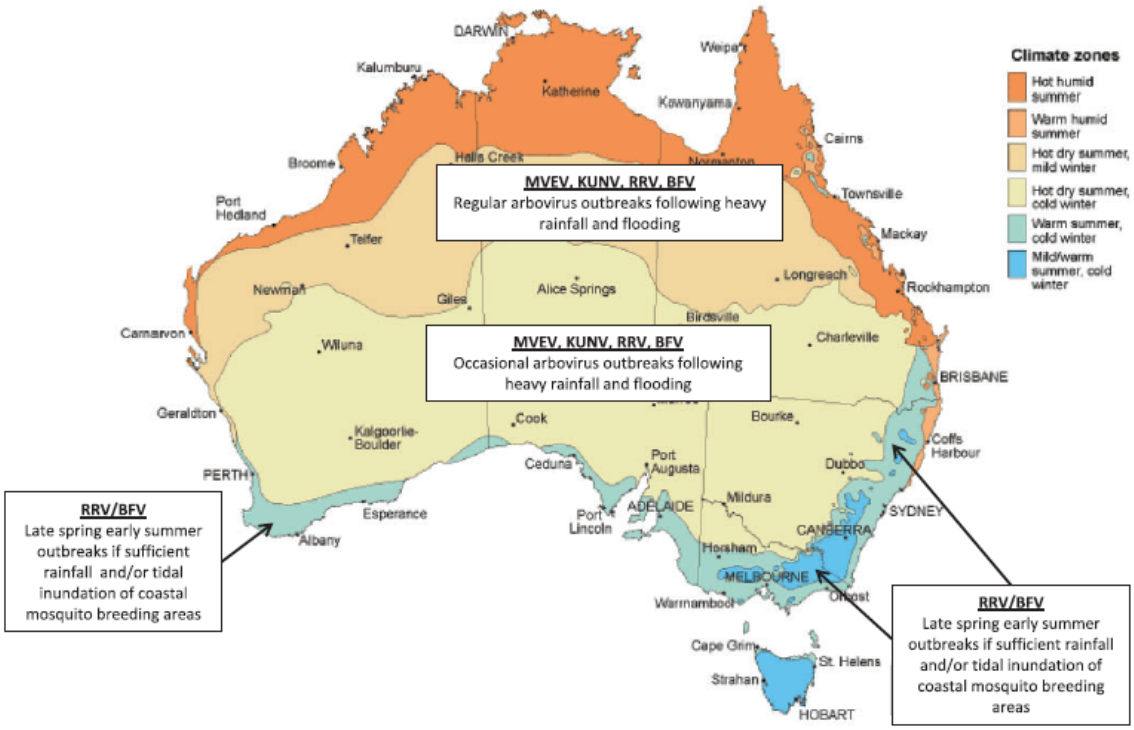


Figure 1.3: Australian climatic zones based on temperature and humidity and the most common associated endemic arboviruses. Taken from Smith (2018).

There have been predictions that changes in climate due to global warming will expand the geographic distribution of vector species, leading to emerging arbovirus diseases in new regions (Gould and Higgs, 2009). Investigation into recent expansions of CHIKV and DENV transmission in Europe indicated that while globalisation and increased international travel were the major contributing factors, increasing temperatures have expanded temporal and spatial limits of transmission (Fischer et al., 2013; Liu-Helmersson

et al., 2016). Attempts have been made to model the effect of climate change on the geographic distribution of vector species (Kraemer et al., 2019; Liu-Helmersson et al., 2019; Monaghan et al., 2018).

1.2 Arbovirus surveillance

Arbovirus surveillance is vital for the prevention, early detection, and control of arboviral outbreaks. Surveillance data helps to inform mosquito management decisions and public health messaging, which often are the only defences against arboviral disease in the absence of vaccine or treatment options. A core component of arbovirus surveillance involves monitoring mosquito populations for arboviral activity, as both mosquito abundance and infection rates can be used to estimate the risk of arbovirus transmission in humans, acting as an early warning system (Gu et al., 2004). Climatic information, and human and animal case notifications are also utilised to better understand spatio-temporal arbovirus activity. This 'One Health' approach to surveillance is necessary to address the complex nature of arbovirus transmission dynamics, which involve human, animal and environmental factors (Dente et al., 2019).

In Australia, long-term arbovirus surveillance activities began in the 1950s, prompted by Murray Valley encephalitis epidemics in eastern Australia, which up until the summer of 1950/51 was known as 'Australian X disease' (Kay and Standfast, 1987; Mackenzie and Broom, 1995). At present, annual arbovirus surveillance is conducted by states and territories during the warmer months (November – April) when mosquito and arbovirus activity peaks. The state/territory surveillance programs produce yearly reports, which the National Arbovirus and Malaria Advisory Committee (NAMAC) collates into a single annual report along with additional human and non-human epidemiological data. NAMAC provides technical advice to the Communicable Diseases Network Australia (CDNA) and makes recommendations about arbovirus surveillance and reporting, vector control and management, and the development of national guidelines and response plans (Knope et al., 2019).

1.2.1 Case-based surveillance

Notifiable disease surveillance systems are used to aid in the prevention and control of disease. Notification data varies globally according to which diseases are considered notifiable, what case definitions are used, and what information is collected and shared (Janati et al., 2015). In Australia, notifiable human arboviral infections are reported by the state/territory hospitals and health practitioners via the National Notifiable Diseases Surveillance System (NNDSS) (Knope et al., 2019). These reports alert vector control programs of circulating arboviruses and can be used to help track the source of infections, prevent further cases, and establish a baseline for activity. Currently, only certain alphaviruses and flaviviruses are notifiable in Australia (Table 1.1). While case-based surveillance can provide up-to-date population level data on arboviral activity, it is often just the 'tip of the iceberg'. Arboviral infections are frequently asymptomatic or produce non-specific illness and therefore go unreported (Butt et al., 2016; Lima-Camara, 2016). Another disadvantage of case-based surveillance is that it is not a preventative surveillance system as it relies on active transmission in humans (Ramírez et al., 2018a).

1.2.2 Sentinel animal surveillance

The zoonotic nature of many arboviruses means that certain animals can be used as indicators of arboviral activity, thereby serving as sentinels. Passive sentinel surveillance can include reports of diseased or dead animals, such as in New York in 1999 when thousands of birds died during a WNV outbreak, which helped to identify the causative virus and define its temporal and geographic limits (Eidson et al., 2001). Active, structured sentinel surveillance involves strategically placing sentinel animals near human populations and wild animal reservoir species to serve as an early warning system for enzootic arbovirus activity (Kwan et al., 2010; Selvey et al., 2014b). Blood samples are periodically taken for antibody testing, which can also be used for virus isolation (Campbell and Hore, 1975). Sentinel animals need to be immunologically naïve so they can undergo seroconversion and ideally develop low or no viremia so as not to serve as reservoir hosts (van den Hurk et al., 2012). A variety of animals have been used as sentinels for arbovirus surveillance worldwide (a comprehensive list can be found in Ramírez et al., 2018). In Australia, sentinel cattle are used for the detection of bovine ephemeral fever virus (BEFV), bluetongue virus (BTV) and Akabane virus (AKAV), and

chickens for MVEV and KUNV (George, 1980; Knope et al., 2019). Despite its popularity as a surveillance tool, sentinel animal surveillance has several limitations. Firstly, it can take weeks for antibody levels to rise after first exposure to an arbovirus thereby delaying detection (Kay et al., 1985). Secondly, serological test results can be difficult to interpret due to the cross-reactivity of closely related viruses, such as MVEV, JEV, WNV and KUNV (Calisher et al., 1989). Thirdly, the animals available to be used as sentinels are not always the preferred host for important vector species and so may not appropriately reflect arbovirus activity (Stephenson et al., 2019). Lastly, the ethical implications and costs associated with establishing and maintaining an animal-based sentinel system can be problematic (Healy et al., 2015; van den Hurk et al., 2012).

1.2.3 Mosquito-based surveillance

Due to species-specific vector competency and feeding preferences, mosquito species abundance information can be used to make predictions about arboviral activity. Furthermore, mosquitoes can be screened for the presence of arboviruses, with detections helping to inform disease and vector management decisions. Therefore, seasonal mosquito monitoring is a common approach in arbovirus surveillance (Engler et al., 2013; Knope et al., 2019; Lustig et al., 2016; Ochieng et al., 2013; Oliver et al., 2018).

A wide variety of trapping techniques have been developed to collect mosquito eggs, larvae and adults in different environments (Day, 2016; Kline, 2006). Carbon dioxide-baited light traps (Figure 1.4A) are most commonly used in surveillance programs due to their ability to capture large numbers of adult mosquitoes from a broad range of species (Drago et al., 2012). Targeted species traps also exist, such as the Biogents Sentinel (BGS) trap (Figure 1.4B), which is more effective at capturing urban *Aedes* species with a human host feeding preference (Farajollahi et al., 2009; Williams et al., 2006). Other traps are designed to capture gravid mosquitoes (Figure 1.4C), which are female mosquitoes looking for water to lay their eggs. Most female mosquitoes require a blood meal to produce and lay their eggs, and so gravid mosquitoes are usually blood-fed and more likely to be carrying arboviruses. Blood-fed mosquitoes can also be used for blood meal analysis to better understand host feeding patterns (Stephenson et al., 2019). Mosquito

traps are usually set up overnight and collected in the morning, with the mosquitoes being placed into a chilled container in order to preserve virus infectivity.

Figure 1.4: Different types of mosquito traps. **(A)** Carbon dioxide-baited CDC light trap; **(B)** Biogents Sentinel trap; **(C)** Frommer updraft gravid trap. Taken from Schilke (2017).

To ascertain mosquito species abundance, mosquitoes in trap catches are identified based on their morphology. Morphological identification is time-consuming, requires specialised expertise and can be difficult when dealing with species that are similar in appearance. Consequently, large trap catches are subsampled when mosquito numbers peak, leaving large numbers of mosquitoes untested (Jacups et al., 2008b). DNA barcoding is an alternative identification method which involves amplifying and sequencing a DNA marker, typically the mitochondrial cytochrome oxidase I (COI), and comparing it to a barcode database, such as the Barcode of Life Data System (BOLD), in order to determine the mosquito species (Batovska et al., 2016). By combining this technique with high-throughput sequencing (HTS), DNA barcoding can be used to identify species in large mixed samples and is termed 'metabarcoding'. This approach can significantly upscale surveillance, with a recent study showing metabarcoding is sensitive enough to detect a single mosquito in a pool of 1,000 (Batovska et al., 2018). Furthermore, the single mosquito was infected with RRV, which could also be detected by using the RRV E2 gene as a diagnostic marker, meaning metabarcoding can be used to perform species identification and virus detection simultaneously (Figure 1.5). DNA-based identification depends on accurate and comprehensive barcode databases containing the species present in the trap, and while efforts have been made to establish such a database in Australia (Batovska et al., 2016), more work needs to be done. Furthermore,

improvements to biases seen in species abundance estimates are needed before metabarcoding can be used as a reliable surveillance tool (Elbrecht and Leese, 2015; Piñol et al., 2015).

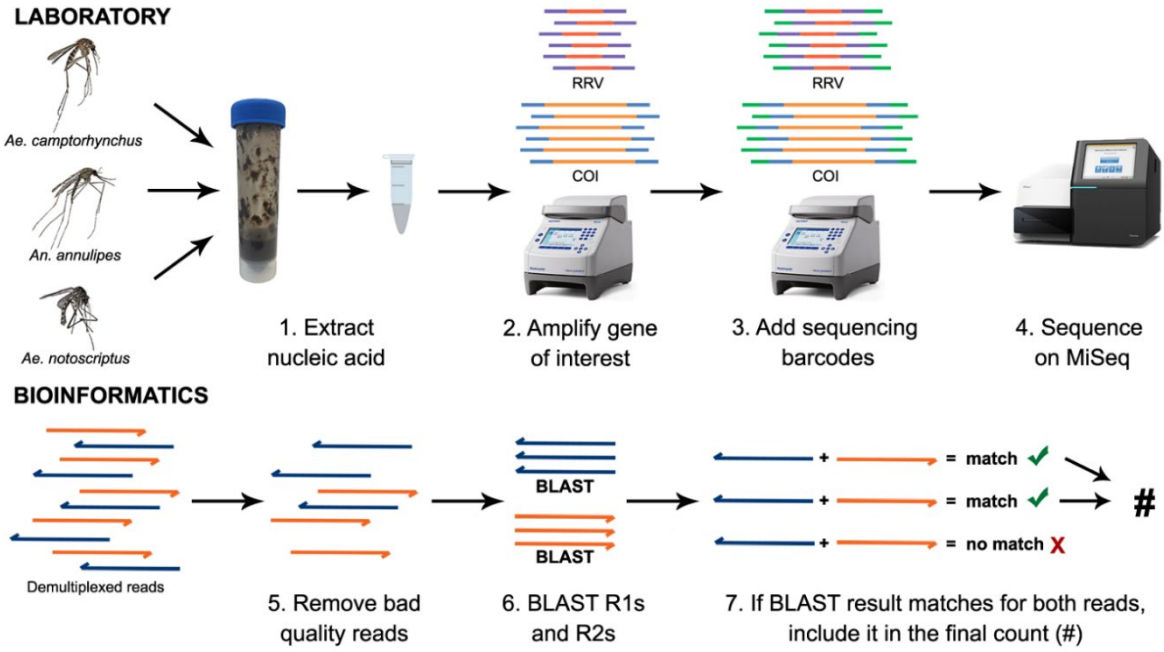


Figure 1.5: The laboratory and bioinformatic process involved in using metabarcoding for species identification and virus detection in a pool of mosquitoes. In this case, COI is being used to identify the mosquito species, and the E2 gene is being used to detect RRV. Taken from Batovska et al. (2018).

1.2.4 Arbovirus screening techniques

Once the mosquitoes have been identified to species they are screened for arboviruses, with virus isolation commonly used by surveillance programs (Knope et al., 2019; Ochieng et al., 2013; Oliver et al., 2018). Virus isolation is an advantageous screening tool as it does not require prior knowledge of the virus, and can be used to phenotypically characterise any novel virus detections (Ramírez et al., 2018a). In the past virus isolation was done using suckling mice, which have subsequently been replaced with cell lines (Temmam et al., 2014). The process involves grouping up to 25 mosquitoes by species, homogenising them and using this to inoculate cell lines that are then observed for cytopathic effect (Russell and Dwyer, 2000). Inoculated cell cultures can also be tested for viral antigen using an enzyme-linked immunosorbent assay (ELISA) or immunofluorescence assay (IFA) with species-specific or genus-reactive monoclonal antibodies (mAbs) (Broom et al.,

1998). In addition to providing virus identification, mAb-based assays can detect viruses that do not have a cytopathic effect on indicator cell lines. A broader spectrum of viruses can be detected by using mAbs that recognise dsRNA produced during viral replication (referred to as MAVRIC – Monoclonal Antibodies against Viral RNA Intermediates in Cell culture) (O'Brien et al., 2015). MAVRIC enables the detection of novel or divergent viruses and has led to the discovery of many new ISVs in Australian mosquitoes (Harrison et al., 2016; McLean et al., 2015; O'Brien et al., 2017, 2020). While useful, cell culture-based approaches in surveillance programs create a bottleneck in the screening process and limit the number of mosquitoes that can be tested. Cell culture also restricts detection to viable viruses capable of growing in the cell lines used (Bexfield and Kellam, 2011).

Over time molecular approaches such as reverse transcription quantitative polymerase chain reaction (RT-qPCR) have gained popularity and are used to supplement or replace virus isolation in arbovirus surveillance programs (Doggett et al., 2019; Engler et al., 2013; Hadler et al., 2015; Knope et al., 2019). RT-qPCR works by extracting and transcribing RNA from a mosquito sample into complementary DNA (cDNA), which is then amplified using DNA polymerase and primers/probes specific to a virus. While specific primers require prior knowledge of virus sequence, targeting RNA means the virus does not have to be viable or culturable, thereby overcoming issues faced by cell culture (Johansen et al., 2002). Furthermore, RT-qPCR can be used to detect multiple arboviruses in a single reaction by either multiplexing primers/probes specific to different viruses (Chao et al., 2007; Ohashi et al., 2004) or by using primers based on conserved gene regions present in an arbovirus genus or family (Pabbaraju et al., 2009; Sánchez-Seco et al., 2001; Scaramozzino et al., 2001). As referred to in the metabarcoding discussion, a molecular approach is sensitive and can be upscaled, with a TaqMan RT-qPCR successfully detecting a single JEV-infected mosquito in pools of up to 5,000 mosquitoes (Ritchie et al., 2003). In Australia, the sensitivity of RT-qPCR has recently led to whole trap grinds (WTGs) becoming the primary arbovirus detection method used in some regions, and involves grinding up the entire trap catch (up to 500 mosquitoes) and testing the supernatant for a variety of arboviruses using RT-qPCR (Doggett et al., 2019).

While RT-qPCR is a useful detection tool implemented in many arbovirus surveillance programs, the cost and equipment involved prohibit its use in low-resource settings. Reverse transcription loop-mediated isothermal amplification (RT-LAMP) is an alternative to RT-qPCR that can amplify nucleic acid in isothermal conditions, removing the need for an expensive thermocycler. RT-LAMP assays can also be designed to produce a colour change or release fluorescence that is observable by eye or with an LED light (Figure 1.6). This technique has been used to detect ZIKV in individual mosquitoes (Lamb et al., 2018; Silva et al., 2019; Yaren et al., 2017), and western equine encephalitis virus (WEEV) and St. Louis encephalitis virus (SLEV) in pools of 50 mosquitoes (Wheeler et al., 2016). The crude lysate from the mosquitoes can be used as input for these assays, eliminating the need for RNA extraction. The ZIKV assay described by Silva et al. (2019) produces a result in 20 minutes, is up to 10,000 times more sensitive than RT-qPCR and costs less than USD $1 per sample. However, RT-LAMP is not always more sensitive than RT-qPCR (Wheeler et al., 2016) and has a more complex primer design, requiring six binding sites instead of two (Yaren et al., 2017).

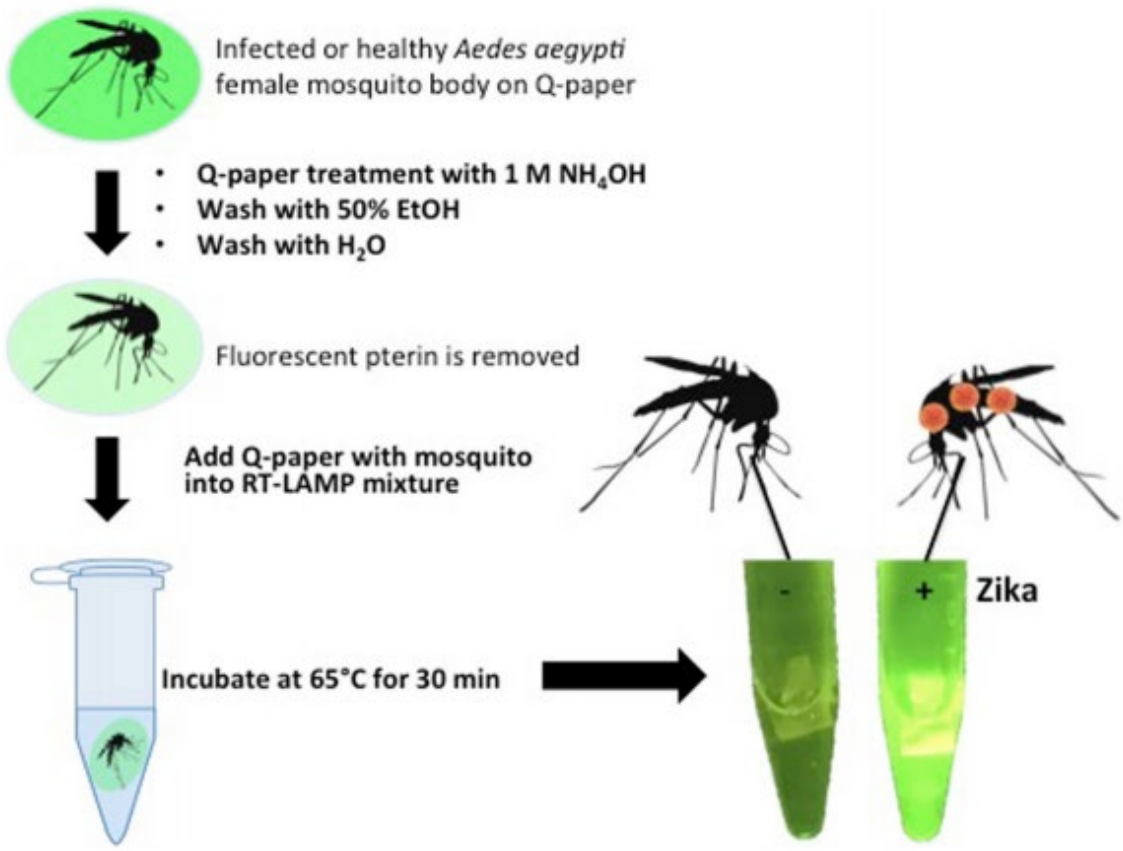


Figure 1.6: An RT-LAMP assay to detect Zika virus in mosquito samples. The fluorescence

produced by the assay is visualised using LED blue light through orange filter glass. Taken from Yaren et al. (2017).

1.2.5 Innovative approaches to arbovirus surveillance

Advances in mosquito-based surveillance have moved from detecting arboviruses directly from mosquitoes to using mosquito secretions, such as saliva and excreta. Female mosquitoes require a blood meal in order to produce eggs, however they also feed on sugar as a source of energy and in the process they expectorate virus from their salivary glands (van den Hurk et al., 2007). This behaviour has been capitalised on with the use of cards baited with honey inside traps, which can then be tested using RT-qPCR to detect any arboviruses that may have been deposited by mosquitoes while feeding. By using Flinders Technology Associates (FTA) cards that are designed to preserve nucleic acids, arboviruses can be detected for up to 28 days post feeding without cold chain transport (Hall-Mendelin et al., 2010). This means that mosquito traps can be deployed for much longer, enabling surveillance in remote locations where collecting traps within 24 hours is not practical. Honey-baited FTA cards have been placed in the commonly used carbon dioxide-baited light traps (Flies et al., 2015), passive box traps (PBT) that are not powered by batteries (van den Hurk et al., 2014), and sentinel mosquito arbovirus capture kits (SMACK) that increase mosquito survival and feeding post-capture (Johnson et al., 2015). This approach to surveillance has shown to be more sensitive than sentinel animals (van den Hurk et al., 2014), however not as sensitive as WTGs (Doggett et al., 2019).

A recent approach to arbovirus detection is the testing of mosquito excreta, which like their saliva, contains arbovirus RNA. RT-qPCR testing of mosquito excreta has been used to detect DENV, KUNV and RRV in laboratory settings (Fontaine et al., 2016; Ramírez et al., 2018b), and KUNV, MVEV and RRV in field settings using modified light and passive mosquito traps (Meyer et al., 2019). Viral RNA is stable in excreta deposited on FTA cards stored at high heat and humidity for 14 days, demonstrating the suitability of this approach for surveillance in remote areas (Ramírez et al., 2019). When compared to saliva, the use of mosquito excreta has been shown to be consistently more sensitive, with 89% vs. 33% samples positive for DENV, 78% vs. 18% for RRV and 68% vs. 11% for KUNV using RT-qPCR (Fontaine et al., 2016; Ramírez et al., 2018b). Unlike saliva, viruses excreted

by mosquitoes do not require an extrinsic incubation period, so viruses that are not vector-borne could also be present, as demonstrated by the detection of hepatitis B virus (HBV) in mosquito excreta using RT-qPCR (Blow et al., 2002).

Like excreta, mosquito blood meals can be used to detect both arboviruses and non-arboviruses, a process termed 'xenosurveillance' by Grubaugh et al. (2015). This approach involves collecting blood-fed mosquitoes from within homes or around areas with lots of human activity, anesthetising them and pressing the blood meals onto FTA cards to use for arbovirus testing (Figure 1.7). RT-qPCR can detect viruses in mosquito blood meals up to 24 hours post-feeding, as well as bacteria and parasites (Fauver et al., 2017). In addition to RT-qPCR, xenosurveillance often involves arbovirus testing via HTS, which allows detection of viruses without prior knowledge of the viral sequence (Fauver et al., 2018; Grubaugh et al., 2015; Yang et al., 2015). Xenosurveillance is appealing as it targets blood-fed mosquitoes, which are more likely to be carrying viruses than the unfed mosquitoes many traps tend to capture. However, collection of blood-fed mosquitoes can be difficult and labour-intensive, thereby limiting the scale of surveillance.

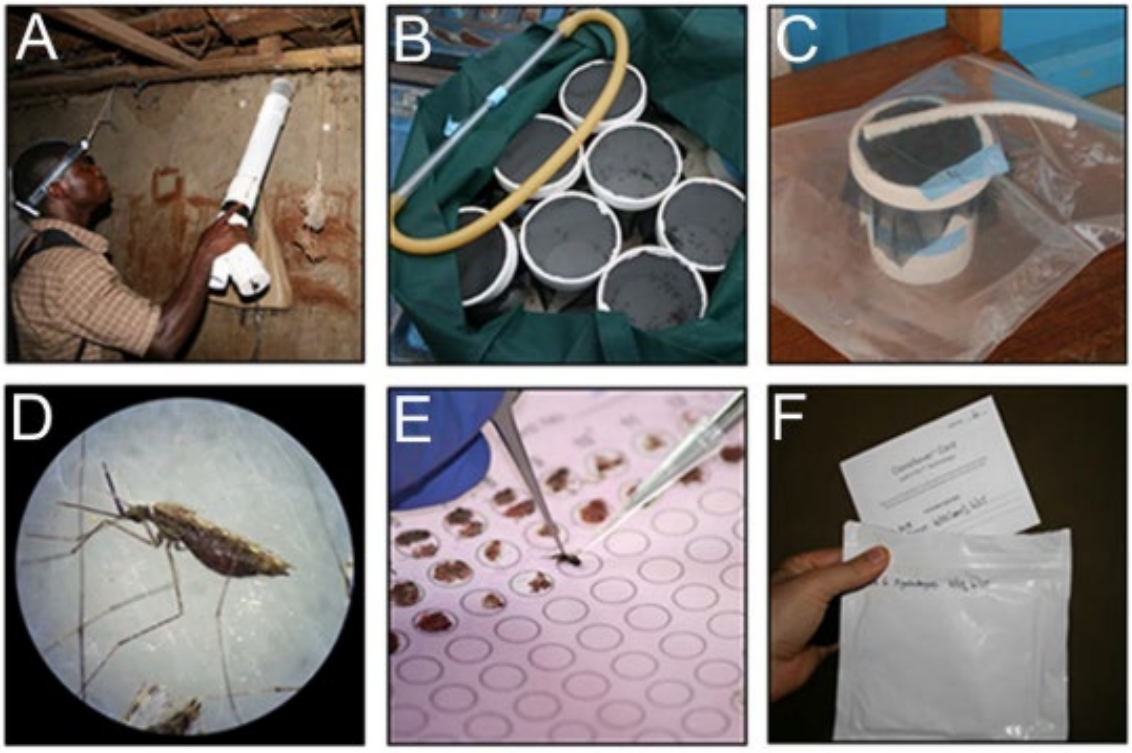


Figure 1.7: The steps involved in a xenosurveillance study performed in Liberia. **(A)** Blood-fed mosquitoes are collected using an aspirator inside homes, **(B)** stored in cartons

according to house, **(C)** anesthetised by triethylamine, and **(D)** identified morphologically. **(E)** The blood meal is applied to an FTA card and **(F)** stored in desiccant pouches for up to three weeks. Taken from Grubaugh et al. (2015).

1.3 Metatranscriptomics for arbovirus surveillance

Metatranscriptomics involves the whole genome, shotgun sequencing of total RNA to create a transcriptional profile of the organisms present in a complex sample, including the viral community (known as the 'virome'). The term metatranscriptomics is sometimes used interchangeably with metagenomics, however the latter can also refer to the sequencing of total or targeted regions of DNA (Siegwald et al., 2017). Metatranscriptomics is a powerful approach to detection as it does not require *a priori* knowledge of the virus, enabling the discovery of thousands of new viruses (Shi et al., 2016; Zhang et al., 2018). This has greatly expanded our understanding of viral diversity, which was previously biased towards disease-causing viruses, detectable by culture-based methods or consensus PCR (Figure 1.8).

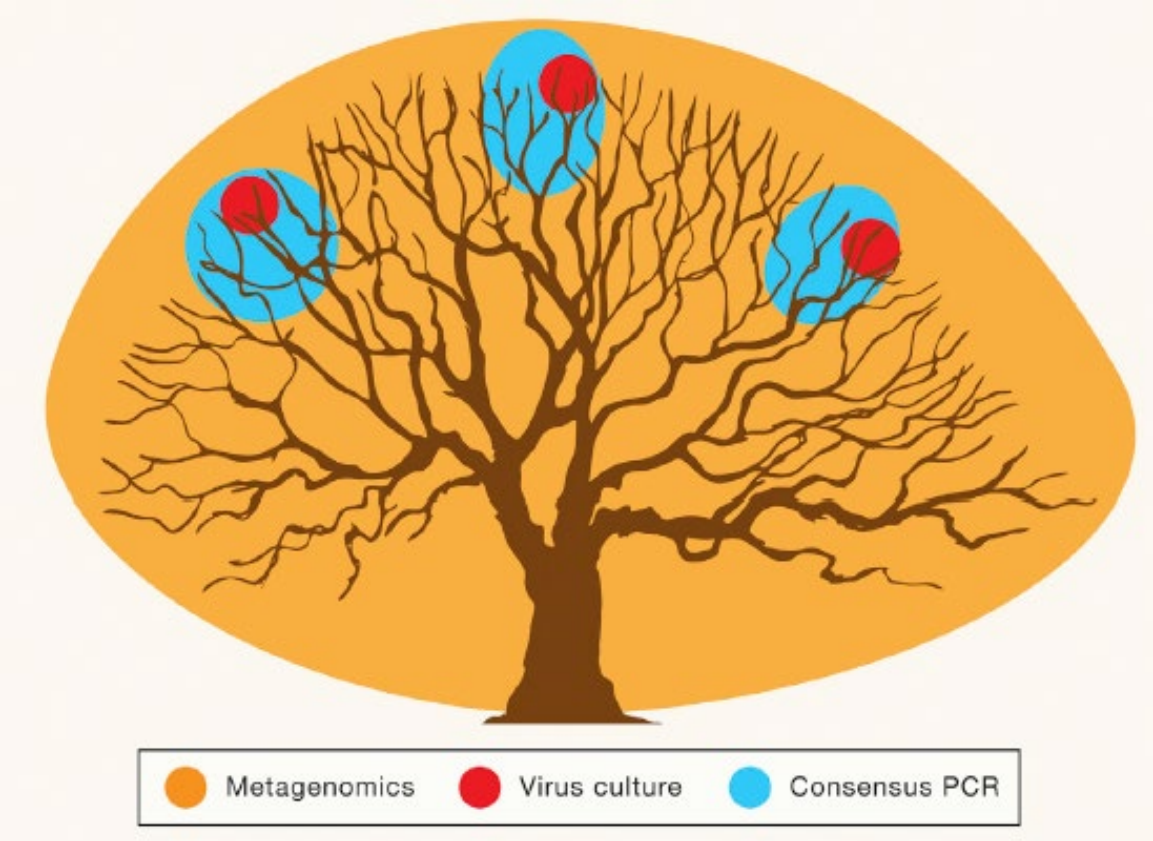


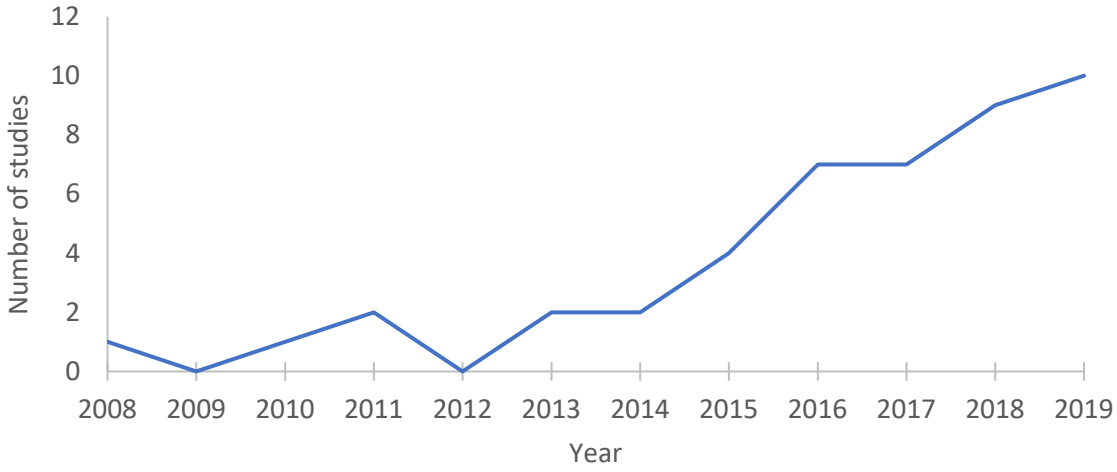
Figure 1.8: A comparison of the ability of metagenomics, cell culture and consensus PCR to discover viral diversity, which is phylogenetically represented as a tree. Taken from Zhang et al. (2018).

The untargeted nature of metatranscriptomics makes it an ideal tool to use for arbovirus surveillance as it enables the detection of new or emerging threats. The use of metatranscriptomics has the potential to significantly upscale surveillance as thousands of mosquitoes can be included in a sample and multiplexed with other samples on a single sequencing run (Du et al., 2020; Xiao et al., 2018a). Unlike RT-qPCR, which targets a specific region of the genome, metatranscriptomics can produce the entire viral genome, which can be used for molecular epidemiology to trace the emergence and transmission of arboviruses in order to inform control and prevention measures (Pollett et al., 2020). The utility of genome-based epidemiology has been demonstrated in arboviral epidemics such as ZIKV (Faria et al., 2017), DENV (Nunes et al., 2014; Tian et al., 2017) and YFV (Faria et al., 2018). Metatranscriptomics offers further utility for surveillance by detecting all of the organisms in a sample, which for mosquitoes includes parasites (Shi et al., 2017), endosymbionts (Chandler et al., 2015; Hall-Mendelin et al., 2013) and the species of the mosquitoes themselves (Batovska et al., 2017; Hall-Mendelin et al., 2013). Despite the continuously decreasing cost of HTS (Hadidi, 2019), metatranscriptomics is still primarily used for research and is not a common feature of arbovirus surveillance programs. A better understanding of the sensitivity and specificity of metatranscriptomic arbovirus detection is required before it can be reliably used for surveillance applications.

While metatranscriptomics is yet to be incorporated as a regular surveillance tool in arbovirus surveillance programs, there are studies that have taken a metatranscriptomic approach to characterise the virome of different mosquito populations, with more studies being conducted globally each year (Figure 1.9). The majority of these studies have explored the diversity of viruses in mosquitoes, with a focus on virus discovery and characterisation (Sadeghi et al., 2018; Shi et al., 2017; Xiao et al., 2018a). In Australia, the first use of metatranscriptomics for arbovirus detection using mosquito samples was by Hall-Mendelin et al. (2013), where they sequenced individual *Aedes aegypti* and *Aedes albopictus* mosquitoes experimentally infected with DENV, YFV or CHIKV. All three

arboviruses were detectable along with other bacterial and fungal organisms, and the species of the mosquitoes were also determined. Colmant et al. (2017) were the first to use metatranscriptomics with field-collected mosquitoes, sequencing two pools of *Anopheles* species trapped in eastern Australia, leading to the discovery of three new viruses in the *Totivirus* and *Orbivirus* genera. This was expanded upon by Shi et al. (2017) when they sequenced 12 Western Australian mosquito populations, revealing 24 species of RNA viruses, 19 of which were newly described. The viral diversity of mosquitoes in south-eastern Australia is yet to be explored using metatranscriptomics.

(A)



(B)

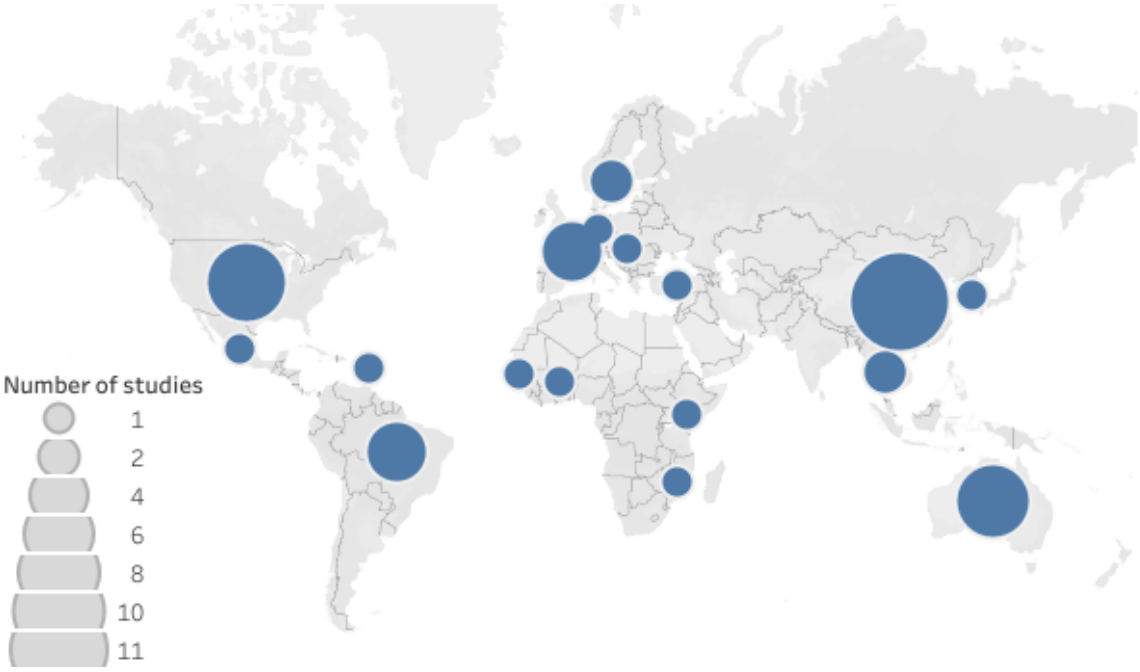


Figure 1.9: Metatranscriptomics for mosquito arbovirus detection in the literature. **(A)** The number of published studies that have used metatranscriptomics to detect arboviruses in mosquito samples each year. **(B)** A map showing the location of the studies with the size of the circles indicating how many were performed in each location. Scopus and PubMed were used on 1 December 2019 to search for published studies that metatranscriptomically sequenced arboviruses from mosquito samples. Studies were only included if the sample was individual or pooled mosquitoes; studies sequencing virus isolates derived from mosquitoes were not included. Tableau v2019.4 (Deardorff, 2016) was used to create the map figure.

1.3.1 Sample preparation

The untargeted nature of metatranscriptomics bestows the ability to detect both known and novel viruses; however, it also results in the majority of sequence reads belonging to the host, reducing sensitivity and cost effectiveness. As much as 92% of the sequences from metatranscriptomic studies were derived from mosquito RNA (Aguiar et al., 2015; Bishop-Lilly et al., 2010), with as little as 0.24% of reads belonging to viruses (Xiao et al., 2018b). The oligo dT mediated enrichment of polyadenylated messenger RNA (mRNA) often used in RNA sequencing experiments is not suitable for broad virus detection due to the existence of non-polyadenylated viruses (Visser et al., 2016). Enriching the viral portion of a sample can be achieved via centrifugation, size filtration and nuclease treatments to remove host cells and nucleic acid (Hall et al., 2014). An example of metatranscriptomic sample preparation involving these methods can be seen in Figure 1.10. Other enrichment approaches involve using sequence-independent amplification (Myrmel et al., 2017), or capture probes targeting a broad range of viral taxa (Briese et al., 2015; Metsky et al., 2019). Virus enrichment can improve detection sensitivity but may introduce biases and lower the accuracy of viral abundance estimates (Halary et al., 2016; Kleiner et al., 2015).

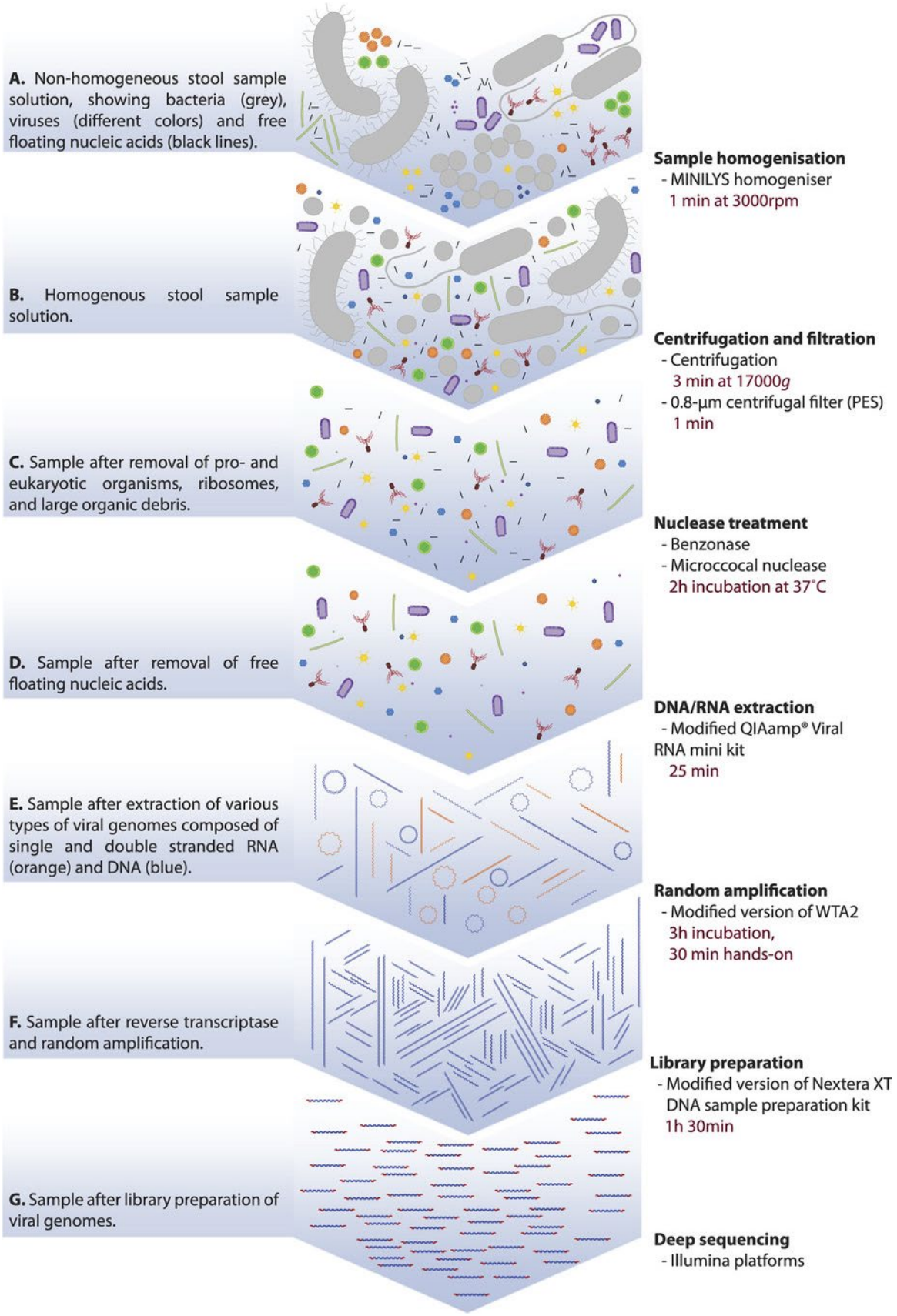

A. Non-homogeneous stool sample solution, showing bacteria (grey), viruses (different colors) and free floating nucleic acids (black lines).
Sample homogenisation
- MINILYS homogeniser
1 min at 3000rpm
B. Homogenous stool sample solution.
Centrifugation and filtration
- Centrifugation
3 min at 17000g
- 0.8-µm centrifugal filter (PES)
1 min
C. Sample after removal of pro- and eukaryotic organisms, ribosomes, and large organic debris.
Nuclease treatment
- Benzonase
- Microccocal nuclease
2h incubation at 37°C
D. Sample after removal of free floating nucleic acids.
DNA/RNA extraction
- Modified QIAamp® Viral RNA mini kit
25 min
E. Sample after extraction of various types of viral genomes composed of single and double stranded RNA (orange) and DNA (blue).
Random amplification
- Modified version of WTA2
3h incubation,
30 min hands-on
F. Sample after reverse transcriptase and random amplification.
Library preparation
- Modified version of Nextera XT DNA sample preparation kit
1h 30min
G. Sample after library preparation of viral genomes.
Deep sequencing
- Illumina platforms

Figure 1.10: An example of a metatranscriptomic protocol designed to increase the proportion of viral reads in a sample. Taken from Conceição-Neto et al. (2015).

An alternative to virus enrichment is the depletion of host ribosomal RNA (rRNA), which constitutes more than 90% of total RNA (Hampton-Marcell et al., 2013). This approach involves using RNA probes specific to host rRNA followed by an RNase H treatment, and is less likely to introduce bias (He et al., 2010; Matranga et al., 2014). Depletion probes in commercially available kits are typically designed for specific human, mouse or bacterial rRNA, limiting their effectiveness with other organisms (Bhagwat et al., 2014; He et al., 2010; Petrova et al., 2017). To overcome this, mosquito-specific rRNA depletion probes have been used in metatranscriptomic studies, increasing viral reads by up to 838.1% (Fauver et al., 2019; Kukutla et al., 2013). Alternatives to probe-based depletion exist, such as using Cas9 with guide RNAs to cleave unwanted sequences, which offers lower cost and input requirements (Gu et al., 2016). Due to genomic variation, any rRNA depletion method will be most effective when target sequences are customised to the mosquito species being used for sample preparation.

1.3.2 Sequencing platforms

A range of different sequencing platforms have been used to perform metatranscriptomic arbovirus detection in mosquito samples (Figure 1.11). These platforms can be broadly categorised into short-read sequencers (second-generation) and long-read sequencers (third-generation), and have been extensively reviewed elsewhere (Depledge et al., 2018; Goodwin et al., 2016). Roche 454, Ion Torrent and Illumina all generate short sequences (700 bp, 400 bp and up to 300 bp, respectively (Goodwin et al., 2016)), and while Roche 454 dominated the earlier metatranscriptomic studies, it has been superseded by Illumina sequencers, which are currently the most popular choice (Figure 1.11). Illumina sequencers include the MiSeq, HiSeq, NextSeq and NovaSeq, and are renowned for their low error rate (<0.1%) and high throughput (Minervini et al., 2020). Illumina has also been the most cost-efficient choice for HTS, however this has been recently challenged by new sequencers from BGI: MGISeq-200, MGISeq-2000 and MGISeq-T7 (Piper et al., 2019).

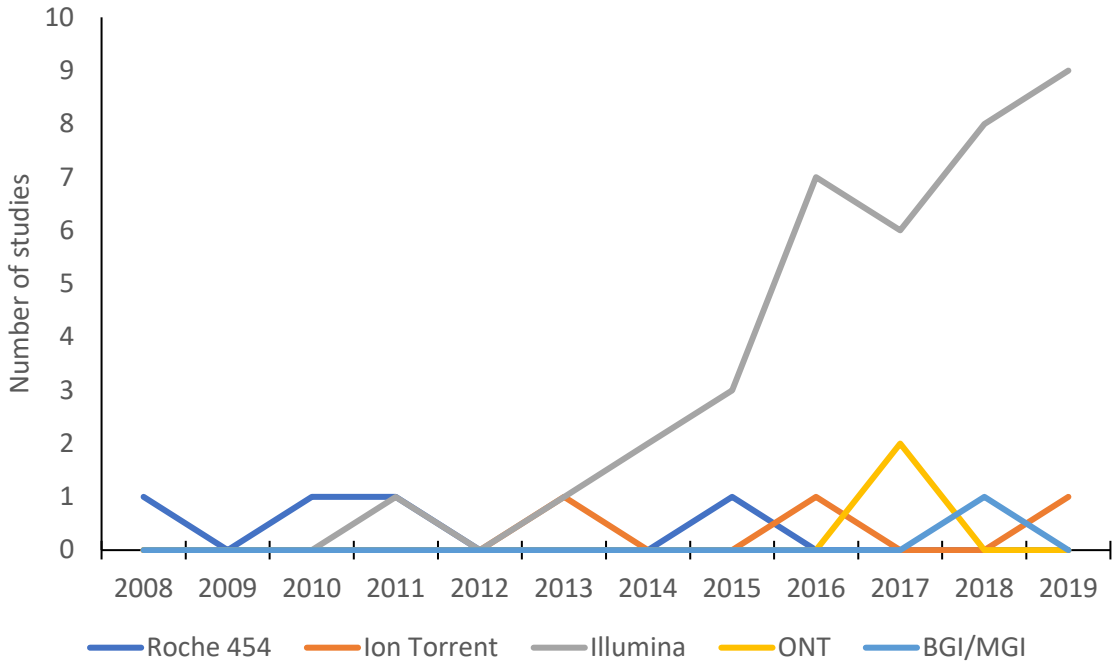


Figure 1.11: The sequencing platforms used in mosquito metatranscriptomic studies. Scopus and PubMed were used on 1 December 2019 to search for published studies that metatranscriptomically sequenced arboviruses from mosquito samples. Studies were only included if the sample was individual or pooled mosquitoes; studies sequencing virus isolates derived from mosquitoes were not included.

Although short-read sequencing has a low error rate and high throughput, the restriction on read length means accurate taxonomic classification is largely dependent on contig assembly, which is complicated by repetitive regions and structural variations. Third-generation sequencers from Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) offer long-read sequencing, with ONT MinION read lengths commonly averaging over 100,000 bp (Branton and Dreamer, 2019) and reaching 2,272,580 bp (Payne et al., 2019). These long reads allow an entire viral genome to be sequenced in a single read (Beaulaurier et al., 2020) and improve recovery of viral genomes from metagenomic samples (Warwick-Dugdale et al., 2019). The main disadvantage of this new technology is a high error rate (up to 5%), but it has been progressively improving (Depledge et al., 2018; Minervini et al., 2020).

The MinION offers unique benefits compared to other sequencers, with its low instrument cost (USD $1000 for sequencer and starter kit), real-time sequencing results, and small

size (Figure 1.12). These features enable real-time, whole-genome pathogen surveillance during outbreaks in resource-limited settings, as demonstrated during the Ebola (Quick et al., 2016), Zika (Faria et al., 2017) and Lassa fever (Kafetzopoulou et al., 2019) outbreaks, the latter of which was performed using a metatranscriptomic approach. To date, the only long-read metatranscriptomic sequencing of arboviruses from mosquitoes has been with the MinION (Figure 1.11). The first study was laboratory-based with a single mosquito specimen (Batovska et al. 2017; Chapter 5), and the second was in-field with a pooled mosquito sample (Russell et al., 2018).

While the MinION instrument costs are low, the cost per run is still high due to the price of flow cells (USD $475–900 each, depending on number purchased), although this is set to change with the recent release of 'Flongle' flow cells, which have less sequencing output but are considerably cheaper (USD $90 each) (Oxford Nanopore Technologies, 2020). A recent feature offered by ONT is direct RNA sequencing, which has been used for whole-genome sequencing of a number of arboviruses (Kim et al., 2019; Wongsurawat et al., 2019), but is yet to be applied using mosquito samples. This unique technology is promising for pathogen surveillance, further streamlining the detection process by obviating the need for cDNA synthesis and enabling the most unbiased approach to RNA sequencing; however, improvements to input requirements, sensitivity and accuracy are needed before it can be used as surveillance tool (Stark et al., 2019).

Figure 1.12: The MinION sequencer. Taken from Oxford Nanopore Technologies (2016).

1.3.3 Bioinformatics

Metatranscriptomic bioinformatics involves analysing millions of sequencing reads in order to determine what viruses are present, which requires significant computational power and staff trained in using a range of specialised computational software. A plethora of software programs exist to identify viral sequences in metatranscriptomic data and these have been reviewed elsewhere (Cantalupo and Pipas, 2019; Nooij et al., 2018; Rose et al., 2016). The typical pipeline involves pre-processing the reads to remove low-quality sequences, removing host reads, assembling the remaining reads into longer, contiguous sequences (contigs), and then comparing the contigs and/or reads to reference databases to determine what viruses are present (Figure 1.13).

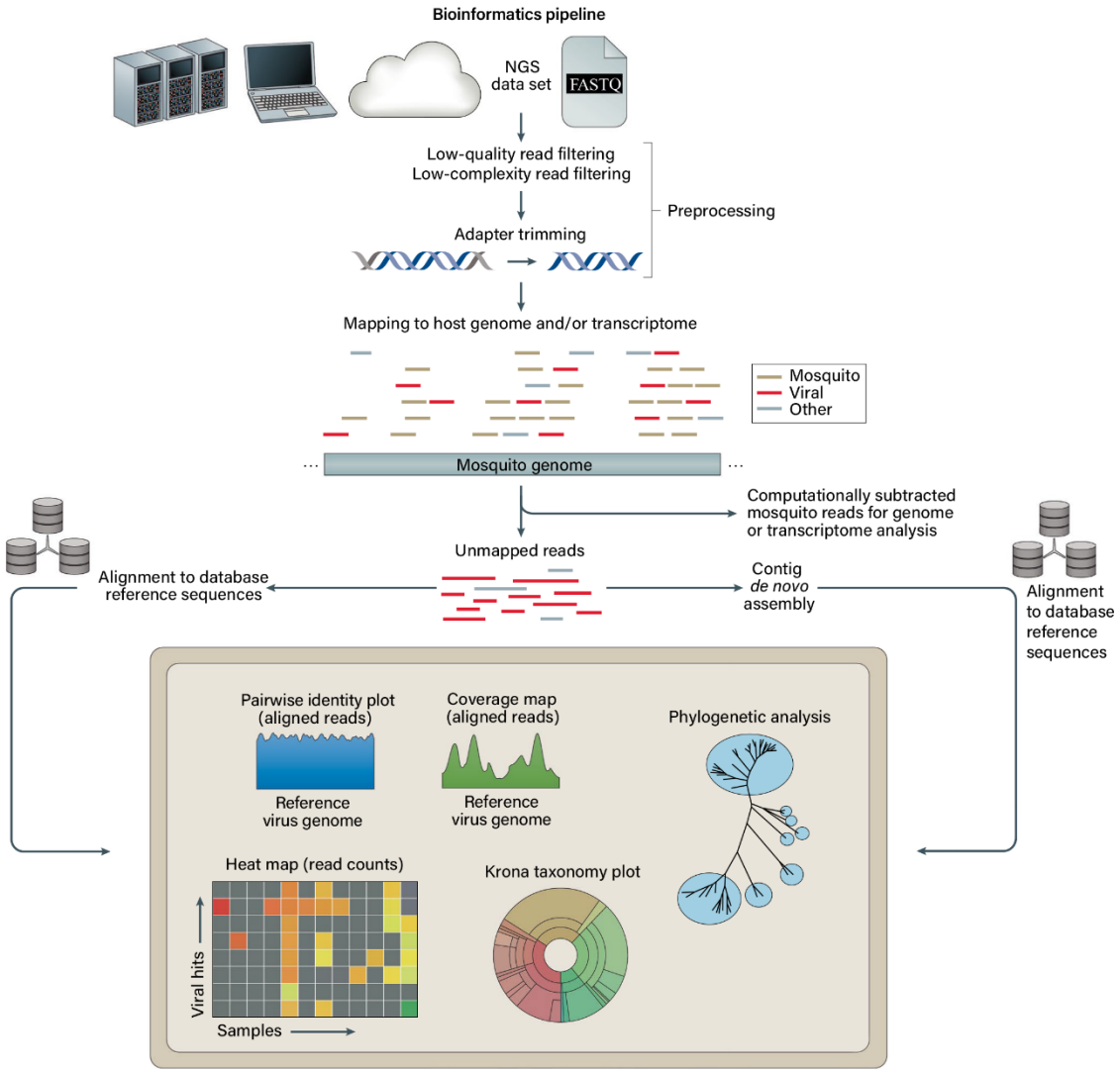


Figure 1.13: An example of a metatranscriptomic data analysis pipeline for virus detection from mosquito samples. Adapted from Chiu and Miller (2019).

Detection of known viruses can be achieved by mapping reads to a reference database using an aligner such as BWA-MEM (Li, 2013) or Bowtie 2 (Langmead and Salzberg, 2012). The alignment can be used to determine how similar the reads are to the reference viral genome via a pairwise identity plot, how much of the genome has been sequenced via a coverage map, and the abundance of the virus relative to others via a heat map of read counts (Figure 1.13). Coverage cut-offs can be established to distinguish positive virus detections from background, with the use of negative controls to determine optimal thresholds (Gu et al., 2019; Schlaberg et al., 2017). Nevertheless, issues with specificity can arise when viruses with sequence homology to viruses in the reference database produce coverage similar to low-level true positive detections, leading to false positive results (Figure 1.14).

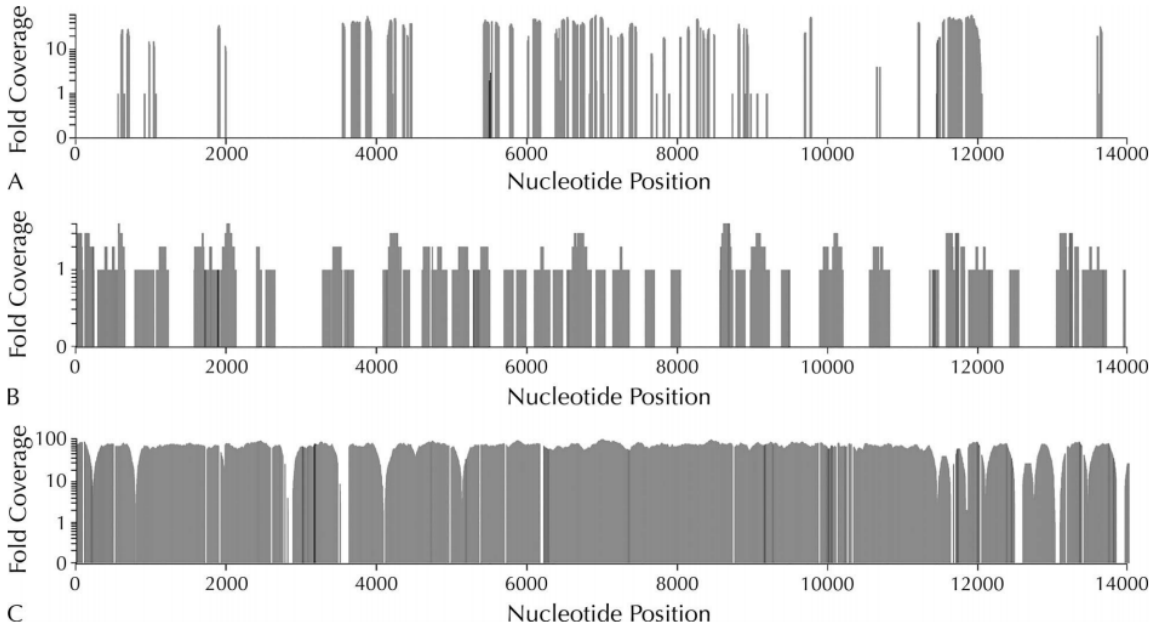


Figure 1.14: Coverage of the measles virus genome produced by a **(A)** false positive detection; **(B)** low-level true positive detection; and **(C)** high-level true positive detection, illustrating the specificity issues that can occur with read-based virus detection. The false positive coverage plot was generated using reads from dolphin morbillivirus, which has partial sequence homology to measles virus. Taken from Schlaberg et al. (2017).

Issues with specificity can be addressed by performing *de novo* assembly to produce contigs, since longer sequences produce more accurate taxonomic classification (McHardy et al., 2007). *De novo* assembly is also required to detect divergent or novel viruses not present in reference databases. There are a wide array of programs available for virome assembly and these have been recently reviewed (Sutton et al., 2019). Popular choices include Trinity (Grabherr et al., 2011), which is designed for transcriptome data, and metaSPAdes (Nurk et al., 2017) or MetaVelvet (Namiki et al., 2012), which are metagenomic extensions of existing assemblers. Assembly quality control can be performed by mapping the reads back to the contigs to ensure coverage, and using tools like TransRate, which checks for chimeric sequences, structural errors and incomplete assembly (Smith-Unna et al., 2016). Good quality contigs can then be compared to a reference database for taxonomic classification using BLASTn, which compares the contig sequences to a nucleotide database, or BLASTx, which translates the contig sequences into six open reading frames (ORFs) and then compares each ORF to a protein database (Camacho et al., 2009). The latter can be very computationally demanding and take weeks or months to run, leading to the development of the program DIAMOND, which can perform the same analysis 20,000 times faster (Buchfink et al., 2015).

Both the read-based and contig-based approaches to virus detection require a reference database. Large, comprehensive databases, like the NCBI nucleotide (nt) database, are powerful as they encompass all publicly available viral sequences, but they also contain errors and can yield inaccurate, unmanageable results (Steinegger and Salzberg, 2020). Targeted, curated databases produce more reliable and practical results, but may also lead to false positive detections and a loss in sensitivity (Marcelino et al., 2020). The correct choice of database often depends on the goal – for instance, if mosquitoes need only to be screened for arboviruses of public health concern, a database containing these genomes can be used. Alternatively, if the aim is to detect emerging, novel pathogens, then a broader and more varied database would be more appropriate (Méric et al., 2019). The use of a reference database for both read-based and contig-based approaches means detection is biased against very divergent genomes. However, the continual addition of viral genomes generated by metatranscriptomic studies to reference databases will help expand known viral diversity and improve detection (Zhang et al., 2018).

Whole genome sequences for viruses also enable better phylogenetic analysis, which can be used to visualise the genetic relationship between different viruses via phylogenetic trees (Zhang et al., 2018). Phylogeny can also be used to classify sequences and help elucidate taxonomy (Linard et al., 2019; Parks et al., 2018; Simmonds and Aiewsakun, 2018). The recent dramatic increase in viral discovery and genome sequences as a result of metatranscriptomic sequencing has challenged existing viral classification, leaving many new viruses unclassified and leading to calls for reorganisation of viral taxonomy (Koonin et al., 2020; Wolf et al., 2018). While the use of metagenomic data has been endorsed by the International Committee on Taxonomy of Viruses (ICTV) (Simmonds et al., 2017), there are concerns about the reclassification of viral taxonomy based on broad phylogenetic comparisons of divergent groups of viruses as it requires the alignment of viral sequences that often share no recognisable similarity, leading to unreliable estimations of evolutionary divergence (Holmes and Duchêne, 2019). Until there is consensus on the correct way to utilise metagenomic data to inform virus taxonomy, many new viruses will be left unclassified (Simmonds and Aiewsakun, 2018).

1.3.4 Contamination

The sensitivity and breadth of virus detection afforded by metatranscriptomic sequencing makes it imperative to manage and control potential contamination issues, which can occur at each stage of the process. During sample preparation, sources of contamination can include the environment (Bukowska-Ośko et al., 2017), humans (Mollerup et al., 2016) and other samples (Eisenhofer et al., 2019; Wyllie et al., 2018). Like with other sensitive molecular techniques, samples should be prepared in a sterile manner with regular decontamination of surfaces and equipment, unidirectional workflow, and physical separation for different assay steps, particularly when exponential amplification is involved (Gu et al., 2019). Regardless, the most common source of contamination during sample preparation are the reagents and kits used (Asplund et al., 2019; Laurence et al., 2014; Naccache et al., 2013; Salter et al., 2014). It is difficult to eliminate reagent-based contamination; therefore, the inclusion of no-template controls (NTCs) at different stages of sample preparation is vital for contaminant detection. A range of software is available

to bioinformatically address contamination (Davis et al., 2018; Schmieder and Edwards, 2011) and cross-contamination (Fiévet et al., 2019).

Contamination can also occur during sequencing when the indexes used to demultiplex samples are misassigned, resulting in reads being incorrectly attributed to other samples. This phenomenon has been referred to in the literature as barcode or index switching, swapping, hopping, cross-talk and misassignment. Index switching can result in up to 10% of reads being misassigned (Sinha et al., 2017), risking false positive detections in surveillance applications. It has been reported across most sequencing platforms, including Illumina (Sinha et al., 2017; Valk et al., 2019), Roche 454 (Carlsen et al., 2012), PacBio (Tedersoo et al., 2018) and ONT MinION (Wick et al., 2018). The rate of index switching can be reduced by using unique dual indexing with sufficiently distinct indices (Costello et al., 2018; MacConaill et al., 2018); limiting the level of free index primers in sequencing libraries by performing thorough clean-up and using Illumina's Free Adaptor Blocking Reagent (Illumina, 2017); avoiding patterned flow cells when using Illumina sequencing (Sinha et al., 2017; Valk et al., 2019); and correcting index switching during data analysis (Larsson et al., 2018). A positive control library made up of foreign (Ji et al., 2020) or synthetic (Hardwick et al., 2018) DNA can be useful for measuring the level of index switching and performing quantitative normalisation between samples.

Due to the many possible sources of contamination, it is important to confirm any detections with an independent method (Gargis et al., 2016). Diagnostic RT-qPCR assays can be used to confirm the presence of an arbovirus in a mosquito sample, or primers can be designed based on the assembled contigs if no assay exists. Sanger sequencing can also be used to confirm metatranscriptomic results and is useful for filling in gaps of newly characterised genomes (Varghese and van Rij, 2018). Ideally, other confirmation methods should also be used, such as virus isolation, electron microscopy or detection of viral coat proteins (Liu et al., 2011). These approaches can also help ascertain if the detection represents an active infection in the mosquito. The method used to confirm detections should be decided before testing starts so the mosquito samples can be preserved appropriately.

1.4 Research overview

1.4.1 Research plan

This PhD project is intended to assess the suitability of metatranscriptomics for arbovirus surveillance. Firstly, mosquito-specific protocols optimised for arbovirus detection will be developed and validated using samples from field traps. The analytical sensitivity and specificity of the metatranscriptomic protocol and associated detection criteria will be established via a spiking experiment using arboviral isolates and a pool of mosquitoes. The established methods will then be evaluated using unsorted, bulk mosquito traps from a variety of locations collected longitudinally as part of an existing arbovirus surveillance program. The utility of metatranscriptomic sequencing data for surveillance will be assessed and compared to existing surveillance approaches. Furthermore, the data will be used to explore the diversity of the mosquito virome in south-eastern Australia. Alternative sequencing approaches will also be investigated to enhance the applicability of metatranscriptomics in biosecurity situations. By developing mosquito-specific protocols and improving our understanding of the sensitivity and specificity of metatranscriptomics, this project will inform the integration of this powerful tool into surveillance programs to enhance arbovirus detection and improve public health outcomes.

1.4.2 Research aim and objectives

Aim: Investigate the use of metatranscriptomics to enhance arbovirus surveillance.

Objectives:

- Develop an optimised metatranscriptomics protocol to detect arboviruses from pools of mosquitoes and assess its analytical sensitivity and specificity.
- Develop criteria based on metatranscriptomic sequencing metrics that can be broadly used to determine if a mosquito pool is positive for an arbovirus.
- Evaluate metatranscriptomic arbovirus detection using unsorted, bulk mosquito traps collected from a variety of locations over time.
- Compare metatranscriptomic arbovirus detection to established surveillance approaches.

- Enhance surveillance capabilities by recovering whole genome sequences from metatranscriptomic data to investigate the genomic epidemiology of arbovirus detections.
- Maximise the utility of metatranscriptomic data by using it to identify mosquito species and explore the broader mosquito virome.
- Establish curated databases of Australian arbovirus and mosquito reference sequences that can be used for future sequencing-based surveillance activities.
- Investigate the use of long-read nanopore sequencing for metatranscriptomic arbovirus detection and compare it to the commonly used short-read Illumina sequencing.

1.5 References

Agarwal, A., Dash, P.K., Singh, A.K., Sharma, S., Gopalan, N., Rao, P.V.L., Parida, M.M., and Reiter, P. (2014). Evidence of experimental vertical transmission of emerging novel ECSA genotype of chikungunya virus in *Aedes aegypti*. PLOS Neglected Tropical Diseases *8*.

Agboli, E., Leggewie, M., Altinli, M., and Schnettler, E. (2019). Mosquito-specific viruses - transmission and interaction. Viruses *11*.

Aguiar, E.R.G.R., Olmo, R.P., Paro, S., Ferreira, F.V., de Faria, I.J. da S., Todjro, Y.M.H., Lobo, F.P., Kroon, E.G., Meignin, C., Gatherer, D., et al. (2015). Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. Nucleic Acids Research *43*, 6191–6206.

Ajamma, Y.U., Onchuru, T.O., Ouso, D.O., Omondi, D., Masiga, D.K., and Villinger, J. (2018). Vertical transmission of naturally occurring Bunyamwera and insect-specific flavivirus infections in mosquitoes from islands and mainland shores of Lakes Victoria and Baringo in Kenya. PLOS Neglected Tropical Diseases *12*, e0006949.

Aldred, J., Campbell, J., Davis, G., Lehmann, N., and Wolstenholme, J. (1990). Barmah Forest virus in the Gippsland Lakes region, Victoria. The Medical Journal of Australia *153*, 434.

Andreadis, T.G., Becnel, J.J., and White, S.E. (2003). Infectivity and pathogenicity of a novel baculovirus, CuniNPV from *Culex nigripalpus* (Diptera: Culicidae) for thirteen species and four genera of mosquitoes. Journal of Medical Entomology *40*, 512–517.

Armstrong, P.M., and Andreadis, T.G. (2010). Eastern equine encephalitis virus in mosquitoes and their role as bridge vectors. Emerging Infectious Diseases *16*, 1869–1874.

Asplund, M., Kjartansdóttir, K.R., Mollerup, S., Vinner, L., Fridholm, H., Herrera, J. a. R., Friis-Nielsen, J., Hansen, T.A., Jensen, R.H., Nielsen, I.B., et al. (2019). Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases *25*, 1277–1285.

Attoui, H., Mohd Jaafar, F., Belhouchet, M., Biagini, P., Cantaloube, J.-F., de Micco, P., and de Lamballerie, X. (2005). Expansion of family *Reoviridae* to include nine-segmented dsRNA viruses: isolation and characterization of a new virus designated Aedes pseudoscutellaris reovirus assigned to a proposed genus (Dinovernavirus). Virology *343*, 212–223.

Azar, S.R., and Weaver, S.C. (2019). Vector competence: what has Zika virus taught us? Viruses *11*, 867.

Ballard, J.W., and Marshall, I.D. (1986). An investigation of the potential of *Aedes camptorhynchus* (Thom.) as a vector of Ross River virus. The Australian Journal of Experimental Biology and Medical Science *64 (Pt 2)*, 197–200.

Barreau, C., Jousset, F.X., and Bergoin, M. (1997). Venereal and vertical transmission of the Aedes albopictus parvovirus in *Aedes aegypti* mosquitoes. The American Journal of Tropical Medicine and Hygiene *57*, 126–131.

Batovska, J., Blacket, M.J., Brown, K., and Lynch, S.E. (2016). Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. Ecology and Evolution *6*, 3001–3011.

Batovska, J., Lynch, S.E., Rodoni, B.C., Sawbridge, T.I., and Cogan, N.O. (2017). Metagenomic arbovirus detection using MinION nanopore sequencing. Journal of Virological Methods *249*, 79–84.

Batovska, J., Lynch, S.E., Cogan, N.O.I., Brown, K., Darbro, J.M., Kho, E.A., and Blacket, M.J. (2018). Effective mosquito and arbovirus surveillance using metabarcoding. Molecular Ecology Resources *18*, 32–40.

Beaulaurier, J., Luo, E., Eppley, J.M., Uyl, P.D., Dai, X., Burger, A., Turner, D.J., Pendleton, M., Juul, S., Harrington, E., et al. (2020). Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. Genome Research gr.251686.119.

Beerntsen, B.T., James, A.A., and Christensen, B.M. (2000). Genetics of mosquito vector competence. Microbiology and Molecular Biology Reviews: MMBR *64*, 115–137.

Belaganahalli, M.N., Maan, S., Maan, N.S., Nomikou, K., Guimera, M., Brownlie, J., Tesh, R., Attoui, H., and Mertens, P.P.C. (2013). Full genome sequencing of Corriparta virus, identifies California mosquito pool virus as a member of the *Corriparta virus* species. PLOS ONE *8*.

Besansky, N.J., Severson, D.W., and Ferdig, M.T. (2003). DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends in Parasitology *19*, 545–546.

Bexfield, N., and Kellam, P. (2011). Metagenomics and the molecular identification of novel viruses. The Veterinary Journal *190*, 191–198.

Bhagwat, A.A., Ying, Z.I., and Smith, A. (2014). Evaluation of ribosomal RNA removal protocols for *Salmonella* RNA-seq projects. Advances in Microbiology *4*.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. Nature *496*, 504–507.

Bishop-Lilly, K.A., Turell, M.J., Willner, K.M., Butani, A., Nolan, N.M.E., Lentz, S.M., Akmal, A., Mateczun, A., Brahmbhatt, T.N., Sozhamannan, S., et al. (2010). Arbovirus detection in insect vectors by rapid, high-throughput pyrosequencing. PLOS Neglected Tropical Diseases *4*, e878.

Blanc, S., and Gutiérrez, S. (2015). The specifics of vector transmission of arboviruses of vertebrates and plants. Current Opinion in Virology *15*, 27–33.

Blasdell, K.R., Voysey, R., Bulach, D., Joubert, D.A., Tesh, R.B., Boyle, D.B., and Walker, P.J. (2012). Kotonkan and Obodhiang viruses: African ephemeroviruses with large and complex genomes. Virology *425*, 143–153.

Blow, J.A., Turell, M.J., Walker, E.D., and Silverman, A.L. (2002). Post-bloodmeal diuretic shedding of hepatitis B virus by mosquitoes (Diptera: Culicidae). Journal of Medical Entomology *39*, 605–612.

Bolling, B.G., Olea-Popelka, F.J., Eisen, L., Moore, C.G., and Blair, C.D. (2012). Transmission dynamics of an insect-specific flavivirus in a naturally infected *Culex pipiens* laboratory colony and effects of co-infection on vector competence for West Nile virus. Virology *427*, 90–97.

Boyd, A.M., and Kay, B.H. (1999). Experimental infection and transmission of Barmah Forest virus by *Aedes vigilax* (Diptera: Culicidae). Journal of Medical Entomology *36*, 186–189.

Boyd, A.M., and Kay, B.H. (2000). Vector competence of *Aedes aegypti, Culex sitiens, Culex annulirostris*, and *Culex quinquefasciatus* (Diptera: Culicidae) for Barmah Forest virus. Journal of Medical Entomology *37*, 660–663.

Branton, D., and Dreamer, D. (2019). Nanopore Sequencing: An Introduction (World Scientific Publishing).

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O.J., and Lipkin, W.I. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio *6*, e01491-01415.

Broom, A.K., Wright, A.E., Mackenzie, J.S., Lindsay, M.D., and Robinson, D. (1989a). Isolation of Murray Valley encephalitis and Ross River viruses from *Aedes normanensis* (Diptera: Culicidae) in Western Australia. Journal of Medical Entomology *26*, 100–103.

Broom, A.K., Mackenzie, J.S., Lindsay, M.D., and Wright, A.E. (1989b). Epidemiology of MVE and Kunjin viruses in Western Australia, 1980-89. In Proceedings of the 5th Symposium, (Brisbane: Commonwealth Scientific and Industrial Research Organisation and Queensland Institute of Medical Research), pp. 14–18.

Broom, A.K., Lindsay, M.D., Wright, A.E., and Mackenzie, J.S. (1993). Arbovirus activity in a remote community in the south-east Kimberley. In Proceedings of the 6th Symposium, (Brisbane: Commonwealth Scientific and Industrial Research Organization and Queensland Institute of Medical Research), pp. 262–266.

Broom, A.K., Lindsay, M.D.A., Johansen, C.A., Wright, A.E. (Tony), and Mackenzie, J.S. (1995). Two possible mechanisms for survival and initiation of Murray Valley encephalitis virus activity in the Kimberley region of Western Australia. The American Journal of Tropical Medicine and Hygiene *53*, 95–99.

Broom, A.K., Hall, R.A., Johansen, C.A., Oliveira, N., Howard, M.A., Lindsay, M.D., Kay, B.H., and Mackenzie, J.S. (1998). Identification of Australian arboviruses in inoculated cell cultures using monoclonal antibodies in ELISA. Pathology *30*, 286–288.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods *12*, 59–60.

Bukowska-Ośko, I., Perlejewski, K., Nakamura, S., Motooka, D., Stokowy, T., Kosińska, J., Popiel, M., Płoski, R., Horban, A., Lipowski, D., et al. (2017). Sensitivity of Next-Generation Sequencing Metagenomic Analysis for Detection of RNA and DNA Viruses in Cerebrospinal Fluid: The Confounding Effect of Background Contamination. In Respiratory Treatment and Prevention, M. Pokorski, ed. (Cham: Springer International Publishing), pp. 53–62.

Butt, A.M., Siddique, S., Gardner, L.M., Sarkar, S., Lancelot, R., and Qamar, R. (2016). Zika virus in Pakistan: the tip of the iceberg? The Lancet Global Health *4*, e913–e914.

Calisher, C.H., and Higgs, S. (2018). The discovery of arthropod-specific viruses in hematophagous arthropods: an open door to understanding the mechanisms of arbovirus and arthropod evolution? Annual Review of Entomology *63*, 87–103.

Calisher, C.H., Karabatsos, N., Dalrymple, J.M., Shope, R.E., Porterfield, J.S., Westaway, E.G., and Brandt, W.E. (1989). Antigenic relationships between flaviviruses as determined by cross-neutralization tests with polyclonal antisera. The Journal of General Virology *70 (Pt 1)*, 37–43.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Campbell, J., and Hore, D.E. (1975). Isolation of Murray Valley encephalitis virus from sentinel chickens. Australian Veterinary Journal *51*, 1–3.

Campbell, G.L., Marfin, A.A., Lanciotti, R.S., and Gubler, D.J. (2002). West Nile virus. The Lancet Infectious Diseases *2*, 519–529.

Cantalupo, P.G., and Pipas, J.M. (2019). Detecting viral sequences in NGS data. Current Opinion in Virology *39*, 41–48.

Carlsen, T., Aas, A.B., Lindner, D., Vrålstad, T., Schumacher, T., and Kauserud, H. (2012). Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? Fungal Ecology *5*, 747–749.

Chandler, J.A., Liu, R.M., and Bennett, S.N. (2015). RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. Frontiers in Microbiology *6*, 185.

Chao, D.-Y., Davis, B.S., and Chang, G.-J.J. (2007). Development of multiplex real-time reverse transcriptase PCR assays for detecting eight medically important flaviviruses in mosquitoes. Journal of Clinical Microbiology *45*, 584–589.

Chiu, C.Y., and Miller, S.A. (2019). Clinical metagenomics. Nature Reviews Genetics *20*, 341–355.

Cholleti, H., Berg, M., Hayer, J., and Blomström, A.-L. (2018). Vector-borne viruses and their detection by viral metagenomics. Infection Ecology & Epidemiology *8*, 1553465.

Ciota, A.T., Chin, P.A., Ehrbar, D.J., Micieli, M.V., Fonseca, D.M., and Kramer, L.D. (2018). Differential effects of temperature and mosquito genetics determine transmissibility of arboviruses by *Aedes aegypti* in Argentina. The American Journal of Tropical Medicine and Hygiene *99*, 417–424.

Colmant, A., Etebari, K., Webb, C., Ritchie, S., Jansen, C., van den Hurk, A., Bielefeldt-Ohmann, H., Hobson-Peters, J., Asgari, S., and Hall, R. (2017). Discovery of new orbiviruses and totivirus from *Anopheles* mosquitoes in eastern Australia. Archives of Virology *162*, 1–6.

Conceição-Neto, N., Zeller, M., Lefrère, H., Bruyn, P.D., Beller, L., Deboutte, W., Yinda, C.K., Lavigne, R., Maes, P., Ranst, M.V., et al. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. Scientific Reports *5*, 16532.

da Costa, C.F., dos Passos, R.A., Lima, J.B.P., Roque, R.A., de Souza Sampaio, V., Campolina, T.B., Secundino, N.F.C., and Pimenta, P.F.P. (2017). Transovarial transmission of DENV in *Aedes aegypti* in the Amazon basin: a local model of xenomonitoring. Parasites & Vectors *10*, 249.

Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T., et al. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics *19*, 332.

Cowled, C., Palacios, G., Melville, L., Weir, R., Walsh, S., Davis, S., Gubala, A., Lipkin, W.I., Briese, T., and Boyle, D. (2009). Genetic and epidemiological characterization of Stretch Lagoon orbivirus, a novel orbivirus isolated from *Culex* and *Aedes* mosquitoes in northern Australia. The Journal of General Virology *90*, 1433–1439.

Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., and Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome *6*, 226.

Day, J.F. (2016). Mosquito oviposition behavior and vector control. Insects *7*, 65.

Deardorff, A. (2016). Tableau (version. 9.1). Journal of the Medical Library Association: JMLA *104*, 182–183.

Dennison, N.J., Jupatanakul, N., and Dimopoulos, G. (2014). The mosquito microbiota influences vector competence for human pathogens. Current Opinion in Insect Science *3*, 6–13.

Dente, M.G., Riccardo, F., Bolici, F., Colella, N.A., Jovanovic, V., Drakulovic, M., Vasic, M., Mamlouk, H., Maazaoui, L., Bejaoui, M., et al. (2019). Implementation of the One Health approach to fight arbovirus infections in the Mediterranean and Black Sea Region: assessing integrated surveillance in Serbia, Tunisia and Georgia. Zoonoses and Public Health *66*, 276–287.

Depledge, D.P., Mohr, I., and Wilson, A.C. (2018). Going the distance: optimizing RNA-seq strategies for transcriptomic analysis of complex viral genomes. Journal of Virology *93*.

Doggett, S., Haniotis, J., Clancy, J., Webb, C., and Toi, C. (2019). The New South Wales Arbovirus Surveillance & Mosquito Monitoring Program. 2018 – 2019 annual report (Westmead: Department of Medical Entomology).

Doherty, R.L., Whitehead, R.H., Wetters, E.J., and Gorman, B.M. (1968). Studies of the epidemiology of arthropod-borne virus infections at Mitchell River Mission, Cape York Peninsula, North Queensland. II. Arbovirus infections of mosquitoes, man and domestic fowls, 1963-1966. Transactions of the Royal Society of Tropical Medicine and Hygiene *62*, 430–438.

Doherty, R.L., Carley, J.G., Kay, B.H., Filippich, C., Marks, E.N., and Frazier, C.L. (1979). Isolation of virus strains from mosquitoes collected in Queensland, 1972–1976. Australian Journal of Experimental Biology and Medical Science *57*, 509–520.

Drago, A., Marini, F., Caputo, B., Coluzzi, M., Torre, A. della, and Pombi, M. (2012). Looking for the gold standard: assessment of the effectiveness of four traps for monitoring mosquitoes in Italy. Journal of Vector Ecology *37*, 117–123.

Du, J., Li, F., Han, Y., Fu, S., Liu, B., Shao, N., Su, H., Zhang, W., Zheng, D., Lei, W., et al. (2020). Characterization of viromes within mosquito species in China. Science China Life Sciences *63*, 1089–1092.

Dutuze, M.F., Nzayirambaho, M., Mores, C.N., and Christofferson, R.C. (2018). A review of Bunyamwera, Batai, and Ngari viruses: understudied orthobunyaviruses with potential one health implications. Frontiers in Veterinary Science *5*.

Eidson, M., Komar, N., Sorhage, F., Nelson, R., Talbot, T., Mostashari, F., and McLean, R. (2001). Crow deaths as a sentinel surveillance system for West Nile virus in the northeastern United States, 1999. Emerging Infectious Diseases *7*, 615–620.

Eisenhofer, R., Minich, J.J., Marotz, C., Cooper, A., Knight, R., and Weyrich, L.S. (2019). Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends in Microbiology *27*, 105–117.

Elbrecht, V., and Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. PLOS ONE *10*, e0130324.

Elena, S.F., and Sanjuán, R. (2005). Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. Journal of Virology *79*, 11555–11558.

Engler, O., Savini, G., Papa, A., Figuerola, J., Groschup, M.H., Kampen, H., Medlock, J., Vaux, A., Wilson, A.J., Werner, D., et al. (2013). European surveillance for West Nile virus in mosquito populations. International Journal of Environmental Research and Public Health *10*, 4869–4895.

Farajollahi, A., Kesavaraju, B., Price, D.C., Williams, G.M., Healy, S.P., Gaugler, R., and Nelder, M.P. (2009). Field efficacy of BG-Sentinel and industry-standard traps for *Aedes albopictus* (Diptera: Culicidae) and West Nile virus surveillance. Journal of Medical Entomology *46*, 919–925.

Faria, N.R., Quick, J., Claro, I.M., Thézé, J., de Jesus, J.G., Giovanetti, M., Kraemer, M.U.G., Hill, S.C., Black, A., da Costa, A.C., et al. (2017). Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature *546*, 406–410.

Faria, N.R., Kraemer, M.U.G., Hill, S.C., Goes de Jesus, J., Aguiar, R.S., Iani, F.C.M., Xavier, J., Quick, J., du Plessis, L., Dellicour, S., et al. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. Science *361*, 894–899.

Fauver, J.R., Gendernalik, A., Weger-Lucarelli, J., Grubaugh, N.D., Brackney, D.E., Foy, B.D., and Ebel, G.D. (2017). The use of xenosurveillance to detect human bacteria, parasites, and viruses in mosquito bloodmeals. The American Journal of Tropical Medicine and Hygiene *97*, 324–329.

Fauver, J.R., Weger-Lucarelli, J., Iii, L.S.F., Bolay, K., Bolay, F.K., Ii, J.W.D., Brackney, D.E., Foy, B.D., Stenglein, M.D., and Ebel, G.D. (2018). Xenosurveillance reflects traditional sampling techniques for the identification of human pathogens: a comparative study in West Africa. PLOS Neglected Tropical Diseases *12*, e0006348.

Fauver, J.R., Akter, S., Morales, A.I.O., Black, W.C., Rodriguez, A.D., Stenglein, M.D., Ebel, G.D., and Weger-Lucarelli, J. (2019). A reverse-transcription/RNase H based protocol for

depletion of mosquito ribosomal RNA facilitates viral intrahost evolution analysis, transcriptomics and pathogen discovery. Virology *528*, 181–197.

Fiévet, A., Bernard, V., Tenreiro, H., Dehainault, C., Girard, E., Deshaies, V., Hupe, P., Delattre, O., Stern, M.-H., Stoppa-Lyonnet, D., et al. (2019). ART-DeCo: easy tool for detection and characterization of cross-contamination of DNA samples in diagnostic next-generation sequencing analysis. European Journal of Human Genetics *27*, 792–800.

Fischer, D., Thomas, S.M., Suk, J.E., Sudre, B., Hess, A., Tjaden, N.B., Beierkuhnlein, C., and Semenza, J.C. (2013). Climate change effects on chikungunya transmission in Europe: geospatial analysis of vector's climatic suitability and virus' temperature requirements. International Journal of Health Geographics *12*, 51.

Flies, E.J., Toi, C., Weinstein, P., Doggett, S.L., and Williams, C.R. (2015). Converting mosquito surveillance to arbovirus surveillance with honey-baited nucleic acid preservation cards. Vector Borne and Zoonotic Diseases *15*, 397–403.

Fontaine, A., Jiolle, D., Moltini-Conclois, I., Lequime, S., and Lambrechts, L. (2016). Excretion of dengue virus RNA by *Aedes aegypti* allows non-destructive monitoring of viral dissemination in individual mosquitoes. Scientific Reports *6*, 1–10.

Gardner, C.L., and Ryman, K.D. (2010). Yellow fever: a reemerging threat. Clinics in Laboratory Medicine *30*, 237–260.

Gargis, A.S., Kalman, L., and Lubin, I.M. (2016). Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. Journal of Clinical Microbiology *54*, 2857–2865.

Gauci, P.J., McAllister, J., Mitchell, I.R., Weir, R.P., Melville, L.F., and Gubala, A.J. (2016). Genomic characterisation of Trubanaman and Gan Gan viruses, two bunyaviruses with potential significance to public health in Australia. Virology Reports *6*, 1–10.

Gaudreault, N.N., Indran, S.V., Balaraman, V., Wilson, W.C., and Richt, J.A. (2019). Molecular aspects of Rift Valley fever virus and the emergence of reassortants. Virus Genes *55*, 1–11.

George, T.D.St. (1980). A sentinel herd system for the study of arbovirus infections in Australia and Papua-New Guinea. Veterinary Research Communications *4*, 39–51.

Go, Y.Y., Balasuriya, U.B.R., and Lee, C. (2014). Zoonotic encephalitides caused by arboviruses: transmission and epidemiology of alphaviruses and flaviviruses. Clinical and Experimental Vaccine Research *3*, 58–77.

Goenaga, S., Kenney, J.L., Duggal, N.K., Delorey, M., Ebel, G.D., Zhang, B., Levis, S.C., Enria, D.A., and Brault, A.C. (2015). Potential for co-infection of a mosquito-specific flavivirus, Nhumirim virus, to block West Nile virus transmission in mosquitoes. Viruses *7*, 5801–5812.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics *17*, 333–351.

Gould, E.A., and Higgs, S. (2009). Impact of climate change and other factors on emerging arbovirus diseases. Transactions of the Royal Society of Tropical Medicine and Hygiene *103*, 109–121.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. Nature Biotechnology *29*, 644–652.

Grubaugh, N.D., Sharma, S., Krajacich, B.J., Iii, L.S.F., Bolay, F.K., Ii, J.W.D., Johnson, W.E., Ebel, G.D., Foy, B.D., and Brackney, D.E. (2015). Xenosurveillance: a novel mosquito-based approach for examining the human-pathogen landscape. PLOS Neglected Tropical Diseases *9*, e0003628.

Grunnill, M., and Boots, M. (2016). How important is vertical transmission of dengue viruses by mosquitoes (Diptera: Culicidae)? Journal of Medical Entomology *53*, 1–19.

Gu, W., Lampman, R., and Novak, R.J. (2004). Assessment of arbovirus vector infection rates using variable size pooling. Medical and Veterinary Entomology *18*, 200–204.

Gu, W., Crawford, E.D., O'Donovan, B.D., Wilson, M.R., Chow, E.D., Retallack, H., and DeRisi, J.L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biology *17*, 41.

Gu, W., Miller, S., and Chiu, C.Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. Annual Review of Pathology *14*, 319–338.

Gulholm, T., and Rawlinson, W.D. (2017). Public health aspects of dengue virus infection relevant to Australia. Microbiology Australia *38*, 191–193.

Gutiérrez-Bugallo, G., Piedra, L.A., Rodriguez, M., Bisset, J.A., Lourenço-de-Oliveira, R., Weaver, S.C., Vasilakis, N., and Vega-Rúa, A. (2019). Vector-borne transmission and evolution of Zika virus. Nature Ecology & Evolution *3*, 561–569.

Gyawali, N., Bradbury, R.S., Aaskov, J.G., and Taylor-Robinson, A.W. (2017). Neglected Australian arboviruses: quam gravis? Microbes and Infection *19*, 388–401.

Hadidi, A. (2019). Next-generation sequencing and CRISPR/Cas13 editing in viroid research and molecular diagnostics. Viruses *11*, 120.

Hadler, J.L., Patel, D., Nasci, R.S., Petersen, L.R., Hughes, J.M., Bradley, K., Etkind, P., Kan, L., and Engel, J. (2015). Assessment of arbovirus surveillance 13 years after introduction of West Nile virus, United States. Emerging Infectious Diseases *21*, 1159–1166.

Halary, S., Temmam, S., Raoult, D., and Desnues, C. (2016). Viral metagenomics: are we missing the giants? Current Opinion in Microbiology *31*, 34–43.

Hall, R.A., and Hobson-Peters, J. (2018). Newly discovered mosquito viruses help control vector-borne viral diseases. Microbiology Australia.

Hall, R.J., Wang, J., Todd, A.K., Bissielo, A.B., Yen, S., Strydom, H., Moore, N.E., Ren, X., Huang, Q.S., Carter, P.E., et al. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. Journal of Virological Methods *195*, 194–204.

Hall-Mendelin, S., Ritchie, S.A., Johansen, C.A., Zborowski, P., Cortis, G., Dandridge, S., Hall, R.A., and van den Hurk, A.F. (2010). Exploiting mosquito sugar feeding to detect mosquito-borne pathogens. PNAS *107*, 11255–11259.

Hall-Mendelin, S., Allcock, R., Kresoje, N., van den Hurk, A.F., and Warrilow, D. (2013). Detection of arboviruses and other micro-organisms in experimentally infected mosquitoes using massively parallel sequencing. PLOS ONE *8*.

Hall-Mendelin, S., McLean, B.J., Bielefeldt-Ohmann, H., Hobson-Peters, J., Hall, R.A., and van den Hurk, A.F. (2016). The insect-specific Palm Creek virus modulates West Nile virus infection in and transmission by Australian mosquitoes. Parasites & Vectors *9*, 414.

Hampton-Marcell, J.T., Moormann, S.M., Owens, S.M., and Gilbert, J.A. (2013). Chapter Nine - Preparation and Metatranscriptomic Analyses of Host–Microbe Systems. In Methods in Enzymology, E.F. DeLong, ed. (Academic Press), pp. 169–185.

Hardwick, S.A., Chen, W.Y., Wong, T., Kanakamedala, B.S., Deveson, I.W., Ongley, S.E., Santini, N.S., Marcellin, E., Smith, M.A., Nielsen, L.K., et al. (2018). Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nature Communications *9*, 1–10.

Harley, D., Sleigh, A., and Ritchie, S. (2001). Ross River virus transmission, infection, and disease: a cross-disciplinary review. Clinical Microbiology Reviews *14*, 909–932.

Harrison, J.J., Warrilow, D., McLean, B.J., Watterson, D., O'Brien, C.A., Colmant, A.M.G., Johansen, C.A., Barnard, R.T., Hall-Mendelin, S., Davis, S.S., et al. (2016). A new orbivirus isolated from mosquitoes in north-western Australia shows antigenic and genetic similarity to Corriparta virus but does not replicate in vertebrate cells. Viruses *8*.

He, S., Wurtzel, O., Singh, K., Froula, J.L., Yilmaz, S., Tringe, S.G., Wang, Z., Chen, F., Lindquist, E.A., Sorek, R., et al. (2010). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nature Methods *7*, 807–812.

Healy, J.M., Reisen, W.K., Kramer, V.L., Fischer, M., Lindsey, N.P., Nasci, R.S., Macedo, P.A., White, G., Takahashi, R., Khang, L., et al. (2015). Comparison of the efficiency and cost of West Nile virus surveillance methods in California. Vector-Borne and Zoonotic Diseases *15*, 147–155.

Hobson-Peters, J., Yam, A.W.Y., Lu, J.W.F., Setoh, Y.X., May, F.J., Kurucz, N., Walsh, S., Prow, N.A., Davis, S.S., Weir, R., et al. (2013). A new insect-specific flavivirus from northern

Australia suppresses replication of West Nile virus and Murray Valley encephalitis virus in co-infected mosquito cells. PLOS ONE *8*, e56534.

Hoffmann, A.A., Montgomery, B.L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P.H., Muzzi, F., Greenfield, M., Durkan, M., Leong, Y.S., Dong, Y., et al. (2011). Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. Nature *476*, 454–457.

Holmes, E.C., and Duchêne, S. (2019). Can sequence phylogenies safely infer the origin of the global virome? MBio *10*.

Hoshino, K., Isawa, H., Tsuda, Y., Yano, K., Sasaki, T., Yuda, M., Takasaki, T., Kobayashi, M., and Sawabe, K. (2007). Genetic characterization of a new insect flavivirus isolated from *Culex pipiens* mosquito in Japan. Virology *359*, 405–414.

Huang, Y., Li, S., Zhao, Q., Pei, G., An, X., Guo, X., Zhou, H., Zhang, Z., Zhang, J., and Tong, Y. (2015). Isolation and characterization of a novel invertebrate iridovirus from adult *Anopheles minimus* (AMIV) in China. Journal of Invertebrate Pathology *127*, 1–5.

van den Hurk, A.F., Nisbet, D.J., Hall, R.A., Kay, B.H., Mackenzie, J.S., and Ritchie, S.A. (2003). Vector competence of Australian mosquitoes (Diptera: Culicidae) for Japanese encephalitis virus. Journal of Medical Entomology *40*, 82–90.

van den Hurk, A.F., Johnson, P.H., Hall-Mendelin, S., Northill, J.A., Simmons, R.J., Jansen, C.C., Frances, S.P., Smith, G.A., and Ritchie, S.A. (2007). Expectoration of flaviviruses during sugar feeding by mosquitoes (Diptera: Culicidae). Journal of Medical Entomology *44*, 845–850.

van den Hurk, A.F., Ritchie, S.A., and Mackenzie, J.S. (2009). Ecology and geographical expansion of Japanese encephalitis virus. Annual Review of Entomology *54*, 17–35.

van den Hurk, A.F., Hall-Mendelin, S., Pyke, A.T., Smith, G.A., and Mackenzie, J.S. (2010). Vector competence of Australian mosquitoes for chikungunya virus. Vector Borne and Zoonotic Diseases (Larchmont, N.Y.) *10*, 489–495.

van den Hurk, A.F., McElroy, K., Pyke, A.T., McGee, C.E., Hall-Mendelin, S., Day, A., Ryan, P.A., Ritchie, S.A., Vanlandingham, D.L., and Higgs, S. (2011). Vector competence of Australian mosquitoes for yellow fever virus. The American Journal of Tropical Medicine and Hygiene *85*, 446–451.

van den Hurk, A.F., Hall-Mendelin, S., Johansen, C.A., Warrilow, D., and Ritchie, S.A. (2012). Evolution of mosquito-based arbovirus surveillance systems in Australia. Journal of Biomedicine & Biotechnology *2012*, 325659.

van den Hurk, A.F., Hall-Mendelin, S., Townsend, M., Kurucz, N., Edwards, J., Ehlers, G., Rodwell, C., Moore, F.A., McMahon, J.L., Northill, J.A., et al. (2014). Applications of a sugar-based surveillance system to track arboviruses in wild mosquito populations. Vector Borne and Zoonotic Diseases (Larchmont, N.Y.) *14*, 66–73.

Illumina (2017). Effects of index misassignment on multiplexing and downstream analysis. [White paper]. Retrieved November 25, 2019 from https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf.

Jacups, S.P., Whelan, P.I., and Currie, B.J. (2008a). Ross River virus and Barmah Forest virus infections: a review of history, ecology, and predictive models, with implications for tropical northern Australia. Vector Borne and Zoonotic Diseases (Larchmont, N.Y.) *8*, 283–297.

Jacups, S.P., Whelan, P.I., Markey, P.G., Cleland, S.J., Williamson, G.J., and Currie, B.J. (2008b). Predictive indicators for Ross River virus infection in the Darwin area of tropical northern Australia, using long-term mosquito trapping data. Tropical Medicine & International Health *13*, 943–952.

Janati, A., Hosseiny, M., Gouya, M.M., Moradi, G., and Ghaderi, E. (2015). Communicable disease reporting systems in the world: a systematic review article. Iranian Journal of Public Health *44*, 1453–1465.

Ji, Y., Huotari, T., Roslin, T., Schmidt, N.M., Wang, J., Yu, D.W., and Ovaskainen, O. (2020). SPIKEPIPE: a metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. Molecular Ecology Resources *20*, 256–267.

Johansen, C.A., Hall, R.A., van den Hurk, A.F., Ritchie, S.A., and Mackenzie, J.S. (2002). Detection and stability of Japanese encephalitis virus RNA and virus viability in dead infected mosquitoes under different storage conditions. The American Journal of Tropical Medicine and Hygiene *67*, 656–661.

Johnson, B.J., Kerlin, T., Hall-Mendelin, S., van den Hurk, A.F., Cortis, G., Doggett, S.L., Toi, C., Fall, K., McMahon, J.L., Townsend, M., et al. (2015). Development and field evaluation of the sentinel mosquito arbovirus capture kit (SMACK). Parasites & Vectors *8*, 509.

Kafetzopoulou, L.E., Pullan, S.T., Lemey, P., Suchard, M.A., Ehichioya, D.U., Pahlmann, M., Thielebein, A., Hinzmann, J., Oestereich, L., Wozniak, D.M., et al. (2019). Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. Science *363*, 74–77.

Kay, B.H. (1979). Seasonal abundance of *Culex annulirostris* and other mosquitoes at Kowanyama, north Queensland, and Charleville, south west Queensland. The Australian Journal of Experimental Biology and Medical Science *57*, 497–508.

Kay, B.H., and Standfast, H.A. (1987). Ecology of Arboviruses and Their Vectors in Australia. In Current Topics in Vector Research: Volume 3, K.F. Harris, ed. (New York, NY: Springer), pp. 1–36.

Kay, B.H., Carley, J.G., Fanning, I.D., and Filippich, C. (1979). Quantitative studies of the vector competence of *Aedes aegypti, Culex annulirostris* and other mosquitoes (Diptera:

Culicidae) with Murray Valley encephalitis and other Queensland arboviruses. Journal of Medical Entomology *16*, 59–66.

Kay, B.H., Fanning, I.D., and Carley, J.G. (1984). The vector competence of Australian *Culex annulirostris* with Murray Valley encephalitis and Kunjin viruses. The Australian Journal of Experimental Biology and Medical Science *62 (Pt 5)*, 641–650.

Kay, B.H., Young, P.L., Hall, R.A., and Fanning, I.D. (1985). Experimental infection with Murray Valley encephalitis virus. Pigs, cattle, sheep, dogs, rabbits, macropods and chickens. The Australian Journal of Experimental Biology and Medical Science *63 (Pt 1)*, 109–126.

Kay, B.H., Fanning, I.D., and Mottram, P. (1989). The vector competence of *Culex annulirostris, Aedes sagax* and *Aedes alboannulatus* for Murray Valley encephalitis virus at different temperatures. Medical and Veterinary Entomology *3*, 107–112.

Kenney, J.L., Solberg, O.D., Langevin, S.A., and Brault, A.C. (2014). Characterization of a novel insect-specific flavivirus from Brazil: potential for inhibition of infection of arthropod cells with medically important flaviviruses. Journal of General Virology, *95*, 2796–2808.

Khin, M.M., and Than, K.A. (1983). Transovarial transmission of dengue 2 virus by *Aedes aegypti* in nature. The American Journal of Tropical Medicine and Hygiene *32*, 590–594.

Kim, J.H., Kim, J., Koo, B.S., Oh, H., Hong, J.J., and Hwang, E.S. (2019). Rapid whole-genome sequencing of Zika viruses using direct RNA sequencing. Journal of Bacteriology and Virology *49*, 115.

Kleiner, M., Hooper, L.V., and Duerkop, B.A. (2015). Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. BMC Genomics *16*, 7.

Kline, D.L. (2006). Traps and trapping techniques for adult mosquito control. Journal of the American Mosquito Control Association *22*, 490–496.

Knope, K., Doggett, S.L., Jansen, C.C., Johansen, C.A., Kurucz, N., Feldman, R., Lynch, S.E., Hobby, M.P., Sly, A., Jardine, A., et al. (2019). Arboviral diseases and malaria in Australia, 2014–15: Annual report of the National Arbovirus and Malaria Advisory Committee. Communicable Diseases Intelligence (2018) *43*.

Koonin, E.V., Dolja, V.V., Krupovic, M., Varsani, A., Wolf, Y.I., Yutin, N., Zerbini, F.M., and Kuhn, J.H. (2020). Global organization and proposed megataxonomy of the virus world. Microbiology and Molecular Biology Reviews *84*.

Kraemer, M.U.G., Reiner, R.C., Brady, O.J., Messina, J.P., Gilbert, M., Pigott, D.M., Yi, D., Johnson, K., Earl, L., Marczak, L.B., et al. (2019). Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*. Nature Microbiology *4*, 854–863.

Kukutla, P., Steritz, M., and Xu, J. (2013). Depletion of ribosomal RNA for mosquito gut metagenomic RNA-seq. Journal of Visualized Experiments: JoVE *(74)*, e50093.

Kwan, J.L., Kluh, S., Madon, M.B., Nguyen, D.V., Barker, C.M., and Reisen, W.K. (2010). Sentinel chicken seroconversions track tangential transmission of West Nile virus to humans in the greater Los Angeles area of California. The American Journal of Tropical Medicine and Hygiene *83*, 1137–1145.

LaBeaud, Ad., Bashir, F., and King, C.H. (2011). Measuring the burden of arboviral diseases: the spectrum of morbidity and mortality from four prevalent infections. Population Health Metrics *9*, 1.

Lamb, L.E., Bartolone, S.N., Tree, M.O., Conway, M.J., Rossignol, J., Smith, C.P., and Chancellor, M.B. (2018). Rapid detection of Zika virus in urine samples and infected mosquitos by reverse transcription-loop-mediated isothermal amplification. Scientific Reports *8*, 1–9.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357–359.

Larsson, A.J.M., Stanley, G., Sinha, R., Weissman, I.L., and Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. Nature Methods *15*, 305–307.

Laurence, M., Hatzis, C., and Brash, D.E. (2014). Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. PLOS ONE *9*, e97876.

Lee, W.-S., Webster, J.A., Madzokere, E.T., Stephenson, E.B., and Herrero, L.J. (2019). Mosquito antiviral defense mechanisms: a delicate balance between innate immunity and persistent viral infection. Parasites & Vectors *12*, 165.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-Bio].

Li, C.-X., Shi, M., Tian, J.-H., Lin, X.-D., Kang, Y.-J., Chen, L.-J., Qin, X.-C., Xu, J., Holmes, E.C., and Zhang, Y.-Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. ELife *4*.

Liehne, C.G., Leivers, S., Stanley, N.F., Alpers, M.P., Paul, S., Liehne, P.F., and Chan, K.H. (1976). Ord River arboviruses--isolations from mosquitoes. The Australian Journal of Experimental Biology and Medical Science *54*, 499–504.

Lima-Camara, T.N. (2016). Emerging arboviruses and public health challenges in Brazil. Revista de Saúde Pública *50*, 36.

Linard, B., Swenson, K., and Pardi, F. (2019). Rapid alignment-free phylogenetic identification of metagenomic sequences. Bioinformatics *35*, 3303–3312.

Lindsay, M.D., Johansen, C., D'Ercole, M., Broom, A.K., Wright, A.E., Condon, R., Smith, D.W., and Mackenzie, J.S. (1992). The epidemiology of outbreaks of Ross River virus infection in Western Australia in 1991-1992. In Not Available, (Commonwealth Scientific & Industrial Organization and Queensland Institute of Medical Research), pp. 72–76.

Lindsay, M.D., Broom, A.K., Wright, A.E., Johansen, C.A., and Mackenzie, J.S. (1993). Ross River virus isolations from mosquitoes in arid regions of Western Australia: implication of vertical transmission as a means of persistence of the virus. The American Journal of Tropical Medicine and Hygiene *49*, 686–696.

Ling, J., Smura, T., Lundström, J.O., Pettersson, J.H.-O., Sironen, T., Vapalahti, O., Lundkvist, Å., and Hesson, J.C. (2019). Introduction and dispersal of Sindbis virus from Central Africa to Europe. Journal of Virology *93*.

Liu, H., Li, M.-H., Zhai, Y.-G., Meng, W.-S., Sun, X.-H., Cao, Y.-X., Fu, S.-H., Wang, H.-Y., Xu, L.-H., Tang, Q., et al. (2010). Banna virus, China, 1987–2007. Emerging Infectious Diseases *16*, 514–517.

Liu, S., Vijayendran, D., and Bonning, B.C. (2011). Next generation sequencing technologies for insect virus discovery. Viruses *3*, 1849–1869.

Liu-Helmersson, J., Quam, M., Wilder-Smith, A., Stenlund, H., Ebi, K., Massad, E., and Rocklöv, J. (2016). Climate change and *Aedes* vectors: 21st century projections for dengue transmission in Europe. EBioMedicine *7*, 267–277.

Liu-Helmersson, J., Rocklöv, J., Sewe, M., and Brännström, Å. (2019). Climate change may enable *Aedes aegypti* infestation in major European cities by 2100. Environmental Research *172*, 693–699.

Lo Presti, A., Cella, E., Angeletti, S., and Ciccozzi, M. (2016). Molecular epidemiology, evolution and phylogeny of chikungunya virus: an updating review. Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases *41*, 270–278.

Lustig, Y., Hindiyeh, M., Orshan, L., Weiss, L., Koren, R., Katz-Likvornik, S., Zadka, H., Glatman-freedman, A., Mendelson, E., and Shulman, L.M. (2016). Mosquito surveillance for 15 years reveals high genetic diversity among West Nile viruses in Israel. The Journal of Infectious Diseases *213*, 1107–1114.

Lutomiah, J.J.L., Mwandawiro, C., Magambo, J., and Sang, R.C. (2007). Infection and vertical transmission of Kamiti river virus in laboratory bred *Aedes aegypti* mosquitoes. Journal of Insect Science *7*.

MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M.S., et al. (2018). Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics *19*, 30.

Mackenzie, J.S., and Broom, A.K. (1995). Australian X disease, Murray Valley encephalitis and the French connection. Veterinary Microbiology *46*, 79–90.

Mackenzie, J.S., Lindsay, M.D., Coelen, R.J., Broom, A.K., Hall, R.A., and Smith, D.W. (1994). Arboviruses causing human disease in the Australasian zoogeographic region. Archives of Virology *136*, 447–467.

Marcelino, V.R., Holmes, E.C., and Sorrell, T.C. (2020). The use of taxon-specific reference databases compromises metagenomic classification. BMC Genomics *21*, 184.

Marshall, I.D. (1979). Epidemiology of Murray Valley encephalitis in eastern Australia - patterns of arbovirus activity and strategies of arbovirus survival. In Proceedings of the 2nd Symposium, (Brisbane: Commonwealth Scientific and Industrial Research Organisation and Queensland Institute of Medical Research), pp. 47–53.

Marshall, I.D. (1985). Epidemiology of arboviruses: Barmah Forest project (Canberra, Australian Capital Territory, Australia: Australian National University).

Marshall, I.D., Woodroofe, G.M., and Hirsch, S. (1982). Viruses recovered from mosquitoes and wildlife serum collected in the Murray Valley of south-eastern Australia, February 1974, during an epidemic of encephalitis. Australian Journal of Experimental Biology and Medical Science *60*, 457–470.

Matranga, C.B., Andersen, K.G., Winnicki, S., Busby, M., Gladden, A.D., Tewhey, R., Stremlau, M., Berlin, A., Gire, S.K., England, E., et al. (2014). Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biology *15*, 519.

McHardy, A.C., Martín, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007). Accurate phylogenetic classification of variable-length DNA fragments. Nature Methods *4*, 63–72.

McLean, D.M. (1953). Transmission of Murray Valley encephalitis virus by mosquitoes. Australian Journal of Experimental Biology and Medical Science *31*, 481–490.

McLean, B.J., Hobson-Peters, J., Webb, C.E., Watterson, D., Prow, N.A., Nguyen, H.D., Hall-Mendelin, S., Warrilow, D., Johansen, C.A., Jansen, C.C., et al. (2015). A novel insect-specific flavivirus replicates only in *Aedes*-derived cells and persists at high prevalence in wild *Aedes vigilax* populations in Sydney, Australia. Virology *486*, 272–283.

McManus, T.J., and Marshall, I.D. (1986). The epidemiology of Ross River virus in Tasmania. In Proceedings of the 4th Symposium, (Brisbane: Commonwealth Scientific and Industrial Research Organisation and Queensland Institute of Medical Research), pp. 127–131.

Méric, G., Wick, R.R., Watts, S.C., Holt, K.E., and Inouye, M. (2019). Correcting index databases improves metagenomic studies. BioRxiv 712166.

Metsky, H.C., Siddle, K.J., Gladden-Young, A., Qu, J., Yang, D.K., Brehio, P., Goldfarb, A., Piantadosi, A., Wohl, S., Carter, A., et al. (2019). Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. Nature Biotechnology *37*, 160–168.

Meyer, D.B., Ramirez, A.L., van den Hurk, A.F., Kurucz, N., and Ritchie, S.A. (2019). Development and field evaluation of a system to collect mosquito excreta for the detection of arboviruses. Journal of Medical Entomology *56*, 1116–1121.

Minervini, C.F., Cumbo, C., Orsini, P., Anelli, L., Zagaria, A., Specchia, G., and Albano, F. (2020). Nanopore sequencing in blood diseases: a wide range of opportunities. Frontiers in Genetics *11*.

Mollerup, S., Friis-Nielsen, J., Vinner, L., Hansen, T.A., Richter, S.R., Fridholm, H., Herrera, J.A.R., Lund, O., Brunak, S., Izarzugaza, J.M.G., et al. (2016). *Propionibacterium acnes*: disease-causing agent or common contaminant? Detection in diverse patient samples by next-generation sequencing. Journal of Clinical Microbiology *54*, 980–987.

Monaghan, A.J., Sampson, K.M., Steinhoff, D.F., Ernst, K.C., Ebi, K.L., Jones, B., and Hayden, M.H. (2018). The potential impacts of 21st century climatic and population changes on human exposure to the virus vector mosquito *Aedes aegypti*. Climatic Change *146*, 487–500.

Myrmel, M., Oma, V., Khatri, M., Hansen, H.H., Stokstad, M., Berg, M., and Blomström, A.-L. (2017). Single primer isothermal amplification (SPIA) combined with next generation sequencing provides complete bovine coronavirus genome coverage and higher sequence depth compared to sequence-independent single primer amplification (SISPA). PLOS ONE *12*, e0187780.

Naccache, S.N., Greninger, A.L., Lee, D., Coffey, L.L., Phan, T., Rein-Weston, A., Aronsohn, A., Hackett, J., Delwart, E.L., and Chiu, C.Y. (2013). The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. Journal of Virology *87*, 11966–11977.

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. Nucleic Acids Research *40*, e155.

Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M.P.G. (2018). Overview of virus metagenomic classification methods and their biological applications. Frontiers in Microbiology *9*, 749.

Nunes, M.R.T., Palacios, G., Faria, N.R., Sousa, E.C., Pantoja, J.A., Rodrigues, S.G., Carvalho, V.L., Medeiros, D.B.A., Savji, N., Baele, G., et al. (2014). Air travel is associated with intracontinental spread of dengue virus serotypes 1-3 in Brazil. PLOS Neglected Tropical Diseases *8*, e2769.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Research *27*, 824–834.

O'Brien, C.A., Hobson-Peters, J., Yam, A.W.Y., Colmant, A.M.G., McLean, B.J., Prow, N.A., Watterson, D., Hall-Mendelin, S., Warrilow, D., Ng, M.-L., et al. (2015). Viral RNA intermediates as targets for detection and discovery of novel and emerging mosquito-borne viruses. PLOS Neglected Tropical Diseases *9*, e0003629.

O'Brien, C.A., McLean, B.J., Colmant, A.M.G., Harrison, J.J., Hall-Mendelin, S., van den Hurk, A.F., Johansen, C.A., Watterson, D., Bielefeldt-Ohmann, H., Newton, N.D., et al.

(2017). Discovery and characterisation of Castlerea virus, a new species of *Negevirus* isolated in Australia. Evolutionary Bioinformatics Online *13*, 1–14.

O'Brien, C.A., Pegg, C.L., Nouwens, A.S., Bielefeldt-Ohmann, H., Huang, B., Warrilow, D., Harrison, J.J., Haniotis, J., Schulz, B.L., Paramitha, D., et al. (2020). A unique relative of Rotifer birnavirus isolated from Australian mosquitoes. Viruses *12*, 1056.

Ochieng, C., Lutomiah, J., Makio, A., Koka, H., Chepkorir, E., Yalwala, S., Mutisya, J., Musila, L., Khamadi, S., Richardson, J., et al. (2013). Mosquito-borne arbovirus surveillance at selected sites in diverse ecological zones of Kenya; 2007 – 2012. Virology Journal *10*, 140.

Ohashi, S., Yoshida, K., Yanase, T., Kato, T., and Tsuda, T. (2004). Simultaneous detection of bovine arboviruses using single-tube multiplex reverse transcription-polymerase chain reaction. Journal of Virological Methods *120*, 79–85.

Öhlund, P., Lundén, H., and Blomström, A.-L. (2019). Insect-specific virus evolution and potential effects on vector competence. Virus Genes *55*, 127–137.

Oliver, J., Lukacik, G., Kokas, J., Campbell, S.R., Kramer, L.D., Sherwood, J.A., and Howard, J.J. (2018). Twenty years of surveillance for Eastern equine encephalitis virus in mosquitoes in New York State from 1993 to 2012. Parasites & Vectors *11*, 362.

Oxford Nanopore Technologies (2016). Sequencing RNA with MinION. URL: https://nanoporetech.com/applications/rna Accessed December 12, 2016.

Oxford Nanopore Technologies (2020). Store - Flow Cells. URL: https://store.nanoporetech.com/flowcells.html/ Accessed May 26, 2020.

Pabbaraju, K., Ho, K.C.Y., Wong, S., Fox, J.D., Kaplen, B., Tyler, S., Drebot, M., and Tilley, P.A.G. (2009). Surveillance of mosquito-borne viruses in Alberta using reverse transcription polymerase chain reaction with generic primers. Journal of Medical Entomology *46*, 640–648.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology *36*, 996–1004.

Patterson, E.I., Villinger, J., Muthoni, J.N., Dobel-Ober, L., and Hughes, G.L. (2020). Exploiting insect-specific viruses as a novel strategy to control vector-borne disease. Current Opinion in Insect Science *39*, 50–56.

Payne, A., Holmes, N., Rakyan, V., and Loose, M. (2019). BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. Bioinformatics *35*, 2193–2198.

Petrova, O.E., Garcia-Alcalde, F., Zampaloni, C., and Sauer, K. (2017). Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. Scientific Reports *7*, 41114.

Piñol, J., Mir, G., Gomez-Polo, P., and Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular Ecology Resources *15*, 819–830.

Piper, A.M., Batovska, J., Cogan, N.O.I., Weiss, J., Cunningham, J.P., Rodoni, B.C., and Blacket, M.J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience *8*.

Pollett, S., Fauver, J.R., Maljkovic Berry, I., Melendrez, M., Morrison, A., Gillis, L.D., Johansson, M.A., Jarman, R.G., and Grubaugh, N.D. (2020). Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. The Journal of Infectious Diseases *221*, S308–S318.

Prow, N.A., Edmonds, J.H., Williams, D.T., Setoh, Y.X., Bielefeldt-Ohmann, H., Suen, W.W., Hobson-Peters, J., van den Hurk, A.F., Pyke, A.T., Hall-Mendelin, S., et al. (2016). Virulence and evolution of West Nile virus, Australia, 1960–2012. Emerging Infectious Diseases *22*, 1353–1362.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature *530*, 228–232.

Ramírez, A.L., van den Hurk, A.F., Meyer, D.B., and Ritchie, S.A. (2018a). Searching for the proverbial needle in a haystack: advances in mosquito-borne arbovirus surveillance. Parasites & Vectors *11*, 320.

Ramírez, A.L., Hall-Mendelin, S., Doggett, S.L., Hewitson, G.R., McMahon, J.L., Ritchie, S.A., and van den Hurk, A.F. (2018b). Mosquito excreta: a sample type with many potential applications for the investigation of Ross River virus and West Nile virus ecology. PLOS Neglected Tropical Diseases *12*, e0006771.

Ramírez, A.L., Hall-Mendelin, S., Hewitson, G.R., McMahon, J.L., Staunton, K.M., Ritchie, S.A., and van den Hurk, A.F. (2019). Stability of West Nile virus (Flaviviridae: Flavivirus) RNA in mosquito excreta. Journal of Medical Entomology *56*, 1135–1138.

Ren, X., Hoiczyk, E., and Rasgon, J.L. (2008). Viral paratransgenesis in the malaria vector *Anopheles gambiae*. PLOS Pathogens *4*.

Rezza, G., Chen, R., and Weaver, S.C. (2017). O'nyong-nyong fever: a neglected mosquito-borne viral disease. Pathogens and Global Health *111*, 271–275.

Ritchie, S.A., Pyke, A.T., Smith, G.A., Northill, J.A., Hall, R.A., van den Hurk, A.F., Johansen, C.A., Montgomery, B.L., and Mackenzie, J.S. (2003). Field evaluation of a sentinel mosquito (Diptera: Culicidae) trap system to detect Japanese encephalitis in remote Australia. Journal of Medical Entomology *40*, 249–252.

Romo, H., Kenney, J.L., Blitvich, B.J., and Brault, A.C. (2018). Restriction of Zika virus infection and transmission in *Aedes aegypti* mediated by an insect-specific flavivirus. Emerging Microbes & Infections *7*, 1–13.

Rose, R., Constantinides, B., Tapinos, A., Robertson, D.L., and Prosperi, M. (2016). Challenges in the analysis of viral metagenomes. Virus Evolution *2*, vew022.

Russell, R.C. (1986). Seasonal abundance of mosquitoes in a native forest of the Murray Valley of Victoria, 1979–1985. Australian Journal of Entomology *25*, 235–240.

Russell, R.C. (1998). Mosquito-borne arboviruses in Australia: the current scene and implications of climate change for human health. International Journal for Parasitology *28*, 955–969.

Russell, R.C., and Dwyer, D.E. (2000). Arboviruses associated with human disease in Australia. Microbes and Infection *2*, 1693–1704.

Russell, J.A., Campos, B., Stone, J., Blosser, E.M., Burkett-Cadena, N., and Jacobs, J.L. (2018). Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol. Scientific Reports *8*, 5417.

Russell, R.C., Cloonan, M.J., Wells, P.J., and Vale, T.G. (1991). Mosquito (Diptera: Culicidae) and arbovirus activity on the south coast of New South Wales, Australia, in 1985-1988. Journal of Medical Entomology *28*, 796–804.

Russell, R.C., Otranto, D., and Wall, R.L. (2013). The Encyclopedia of Medical and Veterinary Entomology (CABI).

Ryan, P.A., and Kay, B.H. (1999). Vector competence of mosquitoes (Diptera: Culicidae) from Maroochy Shire, Australia, for Barmah Forest virus. Journal of Medical Entomology *36*, 856–860.

Ryan, P.A., Do, K.A., and Kay, B.H. (2000). Definition of Ross River virus vectors at Maroochy Shire, Australia. Journal of Medical Entomology *37*, 146–152.

Sadeghi, M., Altan, E., Deng, X., Barker, C.M., Fang, Y., Coffey, L.L., and Delwart, E. (2018). Virome of > 12 thousand *Culex* mosquitoes from throughout California. Virology *523*, 74–88.

Saiyasombat, R., Bolling, B.G., Brault, A.C., Bartholomay, L.C., and Blitvich, B.J. (2011). Evidence of efficient transovarial transmission of Culex flavivirus by *Culex pipiens* (Diptera: Culicidae). Journal of Medical Entomology *48*, 1031–1038.

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology *12*, 87.

Sammels, L.M., Lindsay, M.D., Poidinger, M., Coelen, R.J., and Mackenzie, J.S. (1999). Geographic distribution and evolution of Sindbis virus in Australia. Journal of General Virology *80*, 739–748.

Sánchez-Seco, M.P., Rosario, D., Quiroz, E., Guzmán, G., and Tenorio, A. (2001). A generic nested-RT-PCR followed by sequencing for detection and identification of members of the *Alphavirus* genus. Journal of Virological Methods *95*, 153–161.

Scaramozzino, N., Crance, J.-M., Jouan, A., DeBriel, D.A., Stoll, F., and Garin, D. (2001). Comparison of *Flavivirus* universal primer pairs and development of a rapid, highly sensitive heminested reverse transcription-PCR assay for detection of flaviviruses targeted to a conserved region of the NS5 gene sequences. Journal of Clinical Microbiology *39*, 1922–1927.

Schilke, R.E. (2017). Surveillance of mosquitoes for Zika virus in Westchester County, New York. Preprints *2017030211*.

Schlaberg, R., Chiu, C.Y., Miller, S., Procop, G.W., Weinstock, G., Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, and Microbiology Resource Committee of the College of American Pathologists (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Archives of Pathology & Laboratory Medicine *141*, 776–786.

Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLOS ONE *6*, e17288.

Selvey, L.A., Dailey, L., Lindsay, M., Armstrong, P., Tobin, S., Koehler, A.P., Markey, P.G., and Smith, D.W. (2014a). The changing epidemiology of Murray Valley encephalitis in Australia: the 2011 outbreak and a review of the literature. PLOS Neglected Tropical Diseases *8*, e2656.

Selvey, L.A., Johansen, C.A., Broom, A.K., Antão, C., Lindsay, M.D., Mackenzie, J.S., and Smith, D.W. (2014b). Rainfall and sentinel chicken seroconversions predict human cases of Murray Valley encephalitis in the north of Western Australia. BMC Infectious Diseases *14*, 672.

Shapiro, A., Green, T., Rao, S., White, S., Carner, G., Mertens, P.P.C., and Becnel, J.J. (2005). Morphological and molecular characterization of a cypovirus (*Reoviridae*) from the mosquito *Uranotaenia sapphirina* (Diptera: Culicidae). Journal of Virology *79*, 9430–9438.

Shepard, D.S., Undurraga, E.A., Halasa, Y.A., and Stanaway, J.D. (2016). The global economic burden of dengue: a systematic analysis. The Lancet Infectious Diseases *16*, 935–941.

Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., et al. (2016). Redefining the invertebrate RNA virosphere. Nature *540*, 539–543.

Shi, M., Neville, P., Nicholson, J., Eden, J.-S., Imrie, A., and Holmes, E.C. (2017). High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. Journal of Virology *91*, e00680-17.

Siegwald, L., Touzet, H., Lemoine, Y., Hot, D., Audebert, C., and Caboche, S. (2017). Assessment of common and emerging bioinformatics pipelines for targeted metagenomics. PLOS ONE *12*, e0169563.

Silva, S.J.R. da, Paiva, M.H.S., Guedes, D.R.D., Krokovsky, L., Melo, F.L. de, Silva, M.A.L. da, Silva, A. da, Ayres, C.F.J., and Pena, L.J. (2019). Development and validation of reverse transcription loop-mediated isothermal amplification (RT-LAMP) for rapid detection of ZIKV in mosquito samples from Brazil. Scientific Reports *9*, 1–12.

Simmonds, P., and Aiewsakun, P. (2018). Virus classification - where do you draw the line? Archives of Virology *163*, 2037–2046.

Simmonds, P., Adams, M.J., Benkő, M., Breitbart, M., Brister, J.R., Carstens, E.B., Davison, A.J., Delwart, E., Gorbalenya, A.E., Harrach, B., et al. (2017). Consensus statement: virus taxonomy in the age of metagenomics. Nature Reviews. Microbiology *15*, 161–168.

Sinha, R., Stanley, G., Gulati, G.S., Ezran, C., Travaglini, K.J., Wei, E., Chan, C.K.F., Nabhan, A.N., Su, T., Morganti, R.M., et al. (2017). Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. BioRxiv 125724.

Smith, D.W. (2018). Endemic Australian arboviruses of human health significance. Microbiology Australia *39*, 88–90.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J., and Kelly, S. (2016). TransRate: reference free quality assessment of de novo transcriptome assemblies. Genome Research gr.196469.115.

de Souza, W.M., Acrani, G.O., Romeiro, M.F., Júnior, O.R., Tolardo, A.L., de Andrade, A.A.S., da Silva Gonçalves Vianez Júnior, J.L., de Almeida Medeiros, D.B., Nunes, M.R.T., and Figueiredo, L.T.M. (2016). Complete genome sequence of Piry vesiculovirus. Arch Virol *161*, 2325–2328.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. Nature Reviews Genetics *20*, 631–656.

Steinegger, M., and Salzberg, S.L. (2020). Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. Genome Biology *21*, 115.

Stephenson, E.B., Murphy, A.K., Jansen, C.C., Peel, A.J., and McCallum, H. (2019). Interpreting mosquito feeding patterns in Australia through an ecological lens: an analysis of blood meal studies. Parasites & Vectors *12*, 156.

Sutton, T.D.S., Clooney, A.G., Ryan, F.J., Ross, R.P., and Hill, C. (2019). Choice of assembly software has a critical impact on virome characterisation. Microbiome *7*, 12.

Tedersoo, L., Tooming-Klunderud, A., and Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. The New Phytologist *217*, 1370–1385.

Temmam, S., Davoust, B., Berenger, J.-M., Raoult, D., and Desnues, C. (2014). Viral metagenomics on animals as a tool for the detection of zoonoses prior to human infection? International Journal of Molecular Sciences *15*, 10377–10397.

Thangamani, S., Huang, J., Hart, C.E., Guzman, H., and Tesh, R.B. (2016). Vertical transmission of Zika virus in *Aedes aegypti* mosquitoes. The American Journal of Tropical Medicine and Hygiene *95*, 1169–1173.

Tian, H., Sun, Z., Faria, N.R., Yang, J., Cazelles, B., Huang, S., Xu, B., Yang, Q., Pybus, O.G., and Xu, B. (2017). Increasing airline travel may facilitate co-circulation of multiple dengue virus serotypes in Asia. PLOS Neglected Tropical Diseases *11*, e0005694.

Valk, T. van der, Vezzi, F., Ormestad, M., Dalén, L., and Guschanski, K. (2019). Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. Molecular Ecology Resources *20*.

Varghese, F.S., and van Rij, R.P. (2018). Insect Virus Discovery by Metagenomic and Cell Culture-Based Approaches. In Viral Metagenomics: Methods and Protocols, V. Pantaleo, and M. Chiumenti, eds. (New York, NY: Springer), pp. 197–213.

Viennet, E., Knope, K., Faddy, H.M., Williams, C.R., and Harley, D. (2013). Assessing the threat of chikungunya virus emergence in Australia. Communicable Diseases Intelligence Quarterly Report *37*, E136-143.

Visser, M., Bester, R., Burger, J.T., and Maree, H.J. (2016). Next-generation sequencing for virus detection: covering all the bases. Virology Journal *13*, 85.

Waldock, J., Chandra, N.L., Lelieveld, J., Proestos, Y., Michael, E., Christophides, G., and Parham, P.E. (2013). The role of environmental variables on *Aedes albopictus* biology and chikungunya epidemiology. Pathogens and Global Health *107*, 224–241.

Walker, P.J. (2005). Bovine ephemeral fever in Australia and the world. Current Topics in Microbiology and Immunology *292*, 57–80.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A.C., Allen, M.J., Sullivan, M.B., and Temperton, B. (2019). Long-read viral metagenomics captures abundant and microdiverse viral populations and their niche-defining genomic islands. PeerJ *7*, e6800.

Watson, T.M., and Kay, B.H. (1998). Vector competence of *Aedes notoscriptus* (Diptera: Culicidae) for Ross River virus in Queensland, Australia. Journal of Medical Entomology *35*, 104–106.

Watson, T.M., and Kay, B.H. (1999). Vector competence of *Aedes notoscriptus* (Diptera: Culicidae) for Barmah Forest virus and of *Aedes aegypti* (Diptera: Culicidae) for dengue 1–4 viruses in Queensland, Australia. Journal of Medical Entomology *36*, 508–514.

Weaver, S.C., and Barrett, A.D.T. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. Nature Reviews Microbiology *2*, 789–801.

Webb, C., Doggett, S., and Russell, R. (2016). A Guide to Mosquitoes of Australia (Australia: CSIRO Publishing).

Wheeler, S.S., Ball, C.S., Langevin, S.A., Fang, Y., Coffey, L.L., and Meagher, R.J. (2016). Surveillance for western equine encephalitis, St. Louis encephalitis, and West Nile viruses using reverse transcription loop-mediated isothermal amplification. PLOS ONE *11*, e0147962.

Wick, R.R., Judd, L.M., and Holt, K.E. (2018). Deepbinner: demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. PLOS Computational Biology *14*, e1006583.

Williams, C.R., Long, S.A., Russell, R.C., and Ritchie, S.A. (2006). Field efficacy of the BG-Sentinel compared with CDC Backpack Aspirators and CO2-baited EVS traps for collection of adult *Aedes aegypti* in Cairns, Queensland, Australia. Journal of the American Mosquito Control Association *22*, 296–300.

Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V., and Koonin, E.V. (2018). Origins and evolution of the global RNA virome. MBio *9*.

Wongsurawat, T., Jenjaroenpun, P., Taylor, M.K., Lee, J., Tolardo, A.L., Parvathareddy, J., Kandel, S., Wadley, T.D., Kaewnapan, B., Athipanyasilp, N., et al. (2019). Rapid sequencing of multiple RNA viruses in their native form. Frontiers in Microbiology *10*.

Wyllie, D.H., Robinson, E., Peto, T., Crook, D.W., Ajileye, A., Rathod, P., Allen, R., Jarrett, L., Smith, E.G., and Walker, A.S. (2018). Identifying mixed *Mycobacterium tuberculosis* infection and laboratory cross-contamination during mycobacterial sequencing programs. Journal of Clinical Microbiology *56*.

Xiao, P., Li, C., Zhang, Y., Han, J., Guo, X., Xie, L., Tian, M., Li, Y., Wang, M., Liu, H., et al. (2018a). Metagenomic sequencing from mosquitoes in China reveals a variety of insect and human viruses. Frontiers in Cellular and Infection Microbiology *8*.

Xiao, P., Han, J., Zhang, Y., Li, C., Guo, X., Wen, S., Tian, M., Li, Y., Wang, M., Liu, H., et al. (2018b). Metagenomic analysis of *Flaviviridae* in mosquito viromes isolated from Yunnan Province in China reveals genes from dengue and Zika viruses. Frontiers in Cellular and Infection Microbiology *8*.

Yang, Y., Garver, L.S., Bingham, K.M., Hang, J., Jochim, R.C., Davidson, S.A., Richardson, J.H., and Jarman, R.G. (2015). Feasibility of using the mosquito blood meal for rapid and efficient human and animal virus surveillance and discovery. The American Journal of Tropical Medicine and Hygiene *93*, 1377–1382.

Yaren, O., Alto, B.W., Gangodkar, P.V., Ranade, S.R., Patil, K.N., Bradley, K.M., Yang, Z., Phadke, N., and Benner, S.A. (2017). Point of sampling detection of Zika virus within a multiplexed kit capable of detecting dengue and chikungunya. BMC Infectious Diseases *17*.

Zhang, Y.-Z., Shi, M., and Holmes, E.C. (2018). Using metagenomics to characterize an expanding virosphere. Cell *172*, 1168–1172.

Zivcec, M., Scholte, F.E.M., Spiropoulou, C.F., Spengler, J.R., and Bergeron, É. (2016). Molecular insights into Crimean-Congo hemorrhagic fever virus. Viruses *8*, 106.

CHAPTER 2

Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool

2.1 Chapter preface

This chapter establishes optimised methods for the metatranscriptomic detection of arboviruses from bulk mosquito samples. A spiking experiment was used to determine the analytical sensitivity and specificity of metatranscriptomic arbovirus detection and develop criteria for positive detection. The performance of metatranscriptomic arbovirus detection was compared to established PCR-based methods. The methods and information produced in this chapter will assist in the formulation of standardised metatranscriptomic arbovirus detection protocols for incorporation into routine surveillance programs. In Chapter 3, the methods developed in this chapter were used to screen field traps for arboviruses. The metatranscriptomic data produced in Chapter 3 was used to evaluate the positive detection criteria and thresholds established in this chapter.

This chapter is presented in published format, including the supplementary material published separately online.

2.2 Publication details

Title: Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool

Stage of publication: Published

Journal details: Scientific Reports, 2019, volume 9, issue 1, pages 1–13. DOI: 10.1038/s41598-019-55741-3

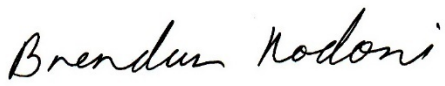
Authors: Jana Batovska, Peter T. Mee, Stacey E. Lynch, Tim I. Sawbridge, and Brendan C. Rodoni

2.3 Statement of joint authorship

All authors contributed to the conception and design of the experiment; PTM and JB prepared the viral spikes; JB performed the nucleic acid extractions, prepared the sequencing libraries, analysed the data, and wrote the manuscript; all authors contributed to the editing of the final manuscript and approved the version submitted for publication.

Statement from co-author confirming the contribution of the PhD candidate:

“As co-author of the manuscript ‘Batovska, J., Mee, P.T., Lynch, S.E., Sawbridge, T.I., and Rodoni, B.C. (2019). Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool. Scientific Reports *9*, 1–13’, I confirm that Jana Batovska has made the contributions listed above.”



Professor Brendan Rodoni

22 June 2020

SCIENTIFIC REPORTS
nature research

OPEN

Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool

Jana Batovska[1,2*], Peter T. Mee[1], Stacey E. Lynch[1*], Tim I. Sawbridge[1,2] & Brendan C. Rodoni[1,2]

The ability to identify all the viruses within a sample makes metatranscriptomic sequencing an attractive tool to screen mosquitoes for arboviruses. Practical application of this technique, however, requires a clear understanding of its analytical sensitivity and specificity. To assess this, five dilutions (1:1, 1:20, 1:400, 1:8,000 and 1:160,000) of Ross River virus (RRV) and Umatilla virus (UMAV) isolates were spiked into subsamples of a pool of 100 *Culex australicus* mosquitoes. The 1:1 dilution represented the viral load of one RRV-infected mosquito in a pool of 100 mosquitoes. The subsamples underwent nucleic acid extraction, mosquito-specific ribosomal RNA depletion, and Illumina HiSeq sequencing. The viral load of the subsamples was also measured using reverse transcription droplet digital PCR (RT-ddPCR) and quantitative PCR (RT-qPCR). Metatranscriptomic sequencing detected both RRV and UMAV in the 1:1, 1:20 and 1:400 subsamples. A high specificity was achieved, with 100% of RRV and 99.6% of UMAV assembled contigs correctly identified. Metatranscriptomic sequencing was not as sensitive as RT-qPCR or RT-ddPCR; however, it recovered whole genome information and detected 19 other viruses, including four first detections for Australia. These findings will assist arbovirus surveillance programs in utilising metatranscriptomics in routine surveillance activities to enhance arbovirus detection.

Metatranscriptomics (total RNA sequencing) enables nontargeted, high-throughput detection and characterisation of viruses in a sample. It can be used to detect both known and novel viruses while providing whole genome information, making it a powerful surveillance tool. Metatranscriptomics has been used in a range of surveillance situations, including detecting viruses in human sewage[1], monitoring viruses in invertebrate vectors such as ticks[2] and vertebrate reservoirs such as bats[3], and tracking virus strains during an outbreak[4]. The successful utilisation of metatranscriptomics in a range of surveillance applications suggests it has potential to enhance current arbovirus (arthropod-borne virus) surveillance programs.

Arboviruses represent a significant burden to human and animal health and include pathogens such as dengue, yellow fever, Zika, chikungunya, bluetongue and equine encephalitis viruses, with dengue virus alone infecting an estimated 390 million people per year[5]. Surveillance programs act as an early warning system for increased transmission risk and enlist tools such as mosquito trapping, virus isolation in cell culture, and targeted molecular virus detection using quantitative PCR (qPCR) assays[6–8]. Metatranscriptomics is a nontargeted method that offers many advantages for arbovirus surveillance programs. It can detect viruses without culturing them, does not require *a priori* knowledge of the viral sequence, has the potential to identify new arboviral threats, elucidates mixed infections, and can provide whole genome or specific protein sequences for molecular epidemiological investigations of outbreaks[9]. Furthermore, it can detect other organisms in a mosquito pool, including endosymbionts such as *Wolbachia*[10], and parasites such as *Leishmania*[11]. The capacity to screen large pools of mosquitoes simultaneously makes metatranscriptomics scalable to adapt to heightened vector abundance[12].

In order to use metatranscriptomics for arbovirus surveillance, the sensitivity and specificity of the method when testing pools of mosquitoes must first be established. A number of studies have used a metatranscriptomic approach to detect viruses in individual mosquitoes using Illumina[10,13], Ion Torrent[14] and Oxford Nanopore[15] sequencing. More often, pools of mosquitoes are sequenced, ranging from five specimens[11] to 6,700 specimens[12]. These studies largely focus on exploring the viral diversity present in various mosquito populations. However, there is a lack of studies looking at gold standard test metrics, such as sensitivity and specificity, of

[1]Agriculture Victoria Research, AgriBio Centre for AgriBioscience, Bundoora, Victoria, Australia. [2]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia. *email: jana.batovska@agriculture.vic.gov.au; stacey.lynch@agriculture.vic.gov.au

metatranscriptomics when testing large pools of mosquitoes for arbovirus surveillance purposes. This is critical when assessing transmission risk and understanding temporal changes in virus abundance. The relationship between viral load and sequencing output needs to be well-defined in order to avoid inaccurate interpretations of sequence data that lead to false positive results (detecting a virus that is not present in the mosquito pool) and false negative results (failing to detect a virus that is present in the mosquito pool).

Laboratory workflows can substantially affect the ability of metatranscriptomic sequencing to detect arboviruses in a mosquito pool. A popular way to increase sensitivity is by enriching for arbovirus using size filtration[16], PEG precipitation[17] or sequence-independent amplification[12]. While this does increase the number of viral sequences, enrichment can also introduce bias[18,19]. An alternate way to increase the number of viral sequences is by depleting the mosquito RNA, generally by targeting highly abundant ribosomal RNA (rRNA). A variety of rRNA depletion kits are available, however, these are not specific to mosquitoes and so custom probes based on mosquito rRNA sequences need to be generated[20,21].

The bioinformatic analyses chosen to process the metatranscriptomic reads can also affect sensitivity and specificity. A common method used to detect viruses in a sample is by mapping reads back to viral reference sequences. However, when dealing with short reads this can lead to false positive results if a virus is present with partial sequence homology to a virus of interest[22]. One way to overcome this problem is by performing *de novo* assembly, where short reads are assembled into longer contiguous sequences (contigs), and then comparing these contigs to a database containing viral reference sequences. This approach can improve specificity because longer fragments are taxonomically classified with greater accuracy[23]. Any viruses detected by the contig-based analysis can then be cross-validated by mapping the sample reads back to the virus reference, which will indicate the breadth and depth of coverage of the virus genome by the reads.

A range of other variables can affect the sensitivity and specificity of metatranscriptomic sequencing including the size and structure (monopartite vs. multipartite) of the virus genome, depth of sequencing, accuracy and completeness of the viral reference database, and the level of host background nucleic acid in the sample[22]. Due to these complications, it can be challenging to establish criteria for positive detection of an arbovirus in a mosquito pool compared to methods like PCR, which is a more targeted detection tool and not impacted by these variables in the same way. As with other detection methods, the use of controls in metatranscriptomics can be used to account for these variables and establish criteria for positive detection. For instance, the addition of a negative control sample that does not contain any viruses can be used to detect viral sequences resulting from physical or cross contamination during the laboratory workflow. Sequence data from the negative control sample can then be used to calculate normalised ratios, for instance the reads per million ratio (RPM-r) where the virus RPM of the sample (RPM_{sample}) is divided by the virus RPM of the negative control (RPM_{neg}). An RPM-r threshold value of 10 has been used to distinguish a true positive detection from contamination for bacteria, fungi and parasites[24].

The purpose of this study is to investigate the analytical sensitivity and specificity of a metatranscriptomic pipeline to detect RNA viruses in mosquito pools for arbovirus surveillance. A spiking experiment was designed in which two viral isolates from distinct RNA viral families (*Togaviridae* and *Reoviridae*) were spiked into clarified subsamples of a pool of 100 mosquitoes (Fig. 1) and sequenced using a library preparation protocol optimised for mosquito samples. The sensitivity and specificity of metatranscriptomic sequencing is assessed and compared with reverse transcription droplet digital PCR (RT-ddPCR) and RT-qPCR. Criteria for positive detection are established, and considerations for laboratory protocol and data analysis are made in an arbovirus surveillance context.

Materials and Methods

Mosquito collection. Adult mosquitoes were collected using carbon dioxide-baited encephalitis virus surveillance traps[25] that were set up overnight and collected the next day. Live mosquitoes were immobilised at −20 °C for 30 minutes and transferred to the laboratory by chilled overnight delivery. Trapping was conducted in November 2016 in Kerang, Victoria, Australia (35.733831 S, 143.925728 E). The mosquitoes were morphologically identified using taxonomic keys[26,27] on top of a cold plate and stored at −20 °C.

Virus spike sample preparation. A pool consisting of 100 *Culex (Culex) australicus* Dobrotworsky & Drummond 1953 (part of the *Culex pipiens* complex) mosquitoes was homogenised in 2 mL of Buffer AVL (Qiagen) using 10 glass beads (3 mm diameter; Sigma-Aldrich) and two 1 minute 1,500 rpm cycles on a 2010 Geno/Grinder (SPEX SamplePrep). The homogenised pool was centrifuged for 5 minutes at 15,344 × *g* and six 120 µL subsamples were taken from the supernatant.

Five of the clarified mosquito subsamples (S1–S5) were spiked with differing dilutions of two cell culture-derived viral isolates (Fig. 1). Ross River virus (RRV) strain T48 (family *Togaviridae*, genus *Alphavirus*)[28] was grown in Vero cells (African green monkey kidney epithelial cells). RRV is a single-stranded, positive-sense RNA virus with a genome approximately 11.8 kb in length[29]. Umatilla virus (UMAV) strain M4941_15 (family *Reoviridae*, genus *Orbivirus*)[30] was grown in C6/36 cells (*Aedes albopictus* cells). UMAV is a double-stranded RNA virus with a 10-segment genome approximately 19.4 kb in length[31]. The viral load of the RRV isolate was 6.9×10^4 copies/ng of RNA, and for the UMAV isolate it was 1.8×10^5 copies/ng of RNA, as measured by RT-ddPCR (see Supplementary Information for details). The S1 clarified mosquito subsample was spiked with 10 µL of the RRV isolate and the S5 subsample was spiked with 10 µL of the UMAV isolate (1:1 spike dilution). This spike represents the viral load of a pool of 100 mosquitoes containing one mosquito infected with RRV as previously described[32], which was assembled and measured by RT-qPCR for comparison (Fig. S1). The RRV and UMAV isolates then underwent a serial 20-fold dilution (1:20; 1:400; 1:8,000; 1:160,000) with 1XTE Buffer pH 8 (Sigma-Aldrich). The remaining clarified mosquito subsamples (S2–S5 for RRV; S4–S1 for UMAV) were spiked with 10 µL of inverse concentrations of the serial dilutions (composition of subsamples seen in Table 1), resulting in 140 µL of input material for the nucleic acid extraction.

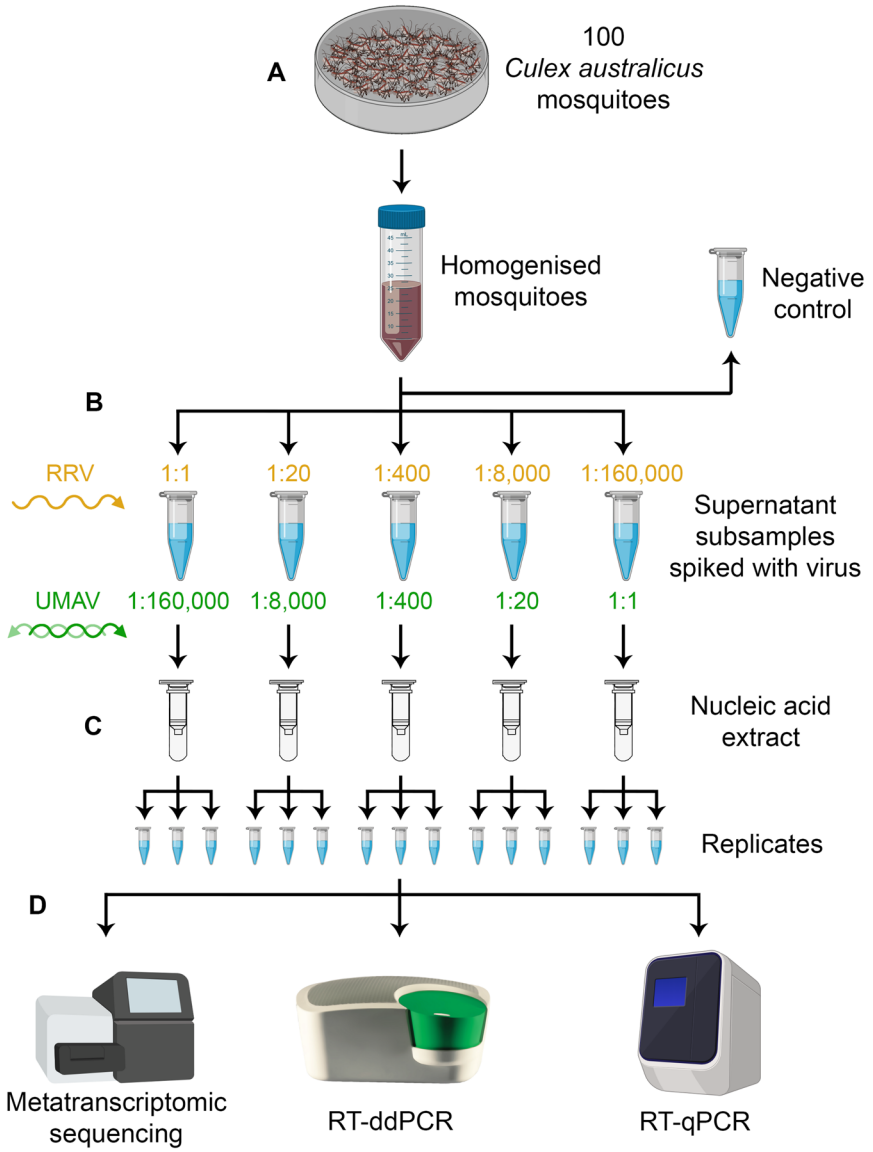


Figure 1. Design of the spiking study. (**A**) 100 mosquitoes were homogenised, centrifuged and the supernatant was subsampled five times, with a sixth subsample taken as a negative control. (**B**) The subsamples were spiked with differing dilutions of Ross River virus (RRV), a monopartite virus, and Umatilla virus (UMAV), a segmented virus. The 1:1 dilution represented the viral load of a single RRV-infected mosquito in a pool of 100. (**C**) Nucleic acid was extracted and split into three technical replicates. (**D**) Viral load was measured using metatranscriptomic sequencing, and reverse transcription droplet digital PCR (RT-ddPCR) and quantitative PCR (RT-qPCR). Created with BioRender.com.

The sixth 120 µL clarified mosquito subsample had 20 µL of 1XTE Buffer pH 8 added to it and was used as a negative control to ensure the mosquito pool was free of both RRV and UMAV, and to account for any contamination and background noise during sequencing.

Nucleic acid extraction. Nucleic acid was extracted from the six clarified mosquito subsamples using the QIAamp Viral RNA Mini Kit (Qiagen) according to manufacturers' instruction, except that carrier RNA was not used. The final elution volume of 80 µL in Buffer AVE was split into three 25 µL aliquots to create technical replicates for each of the spiked clarified mosquito subsamples. This resulted in a total of 15 RNA samples, and one negative control sample (Table 1 and Fig. 1). Due to the double-stranded RNA genome structure of UMAV, all of the RNA was heat-denatured at 100 °C for 1 minute[33] and immediately placed on ice. The RNA was quantified using a Qubit RNA HS Assay Kit (Thermo Fisher Scientific) and then stored at −80 °C until further analysis.

Virus spike sample quantification using metatranscriptomic sequencing. Metatranscriptomic sequencing was performed on all 15 spiked mosquito pool samples and the unspiked negative control sample. Sequencing libraries were prepared using the strand-specific NuGEN Ovation Universal RNA-Seq System with

Sample name	S1.1	S1.2	S1.3	S2.1	S2.2	S2.3	S3.1	S3.2	S3.3	S4.1	S4.2	S4.3	S5.1	S5.2	S5.3	Neg
RRV spike dilution	1:1			1:20			1:400			1:8,000			1:160,000			0
UMAV spike dilution	1:160,000			1:8,000			1:400			1:20			1:1			0
Reads (millions)	22.6	23.45	19.1	19.45	17.1	17.05	18.65	20.45	21.05	22.55	18.9	21.2	23.25	19.1	19.45	21.7
Viral reads (%)	16.0	15.4	15.9	15.9	16.9	16.3	15.2	14.6	15.1	14.6	15.6	11.6	17.1	16.9	16.9	15.1
No. of viral contigs	527	536	545	540	551	494	482	557	539	550	511	518	497	491	513	529
RRV contigs	20	16	26	9	25	14	2	6	4	0	3	0	3	0	1	2
UMAV contigs	4	21	6	16	10	6	64	48	52	32	33	33	38	34	48	6
RRV (%)[a]	100.0	100.0	100.0	100.0	97.0	94.5	3.7	11.8	11.1	0.0	5.1	0.0	5.0	0.0	4.4	3.7
UMAV (%)[b] Total:	5.8	26.0	7.4	21.2	16.2	9.1	92.9	83.0	90.3	98.7	98.8	98.8	98.6	99.1	98.9	7.9
Seg 1 (VP1/RdRp)	0.0	13.8	6.8	7.0	24.4	0.0	82.8	73.3	86.4	100.0	100.0	100.0	100.0	100.0	100.0	0.0
Seg 2 (VP2/T2)	0.0	9.4	0.0	7.3	0.0	8.6	98.0	65.7	93.3	99.0	99.0	98.8	99.0	99.2	99.1	17.4
Seg 3 (VP3)	10.9	29.3	0.0	8.9	9.8	0.0	98.5	93.9	98.6	100.0	100.0	100.0	99.2	100.0	100.0	0.0
Seg 4 (VP4/CaP)	0.0	23.5	0.0	21.7	0.0	0.0	95.3	83.1	94.8	96.1	95.5	96.6	95.4	96.7	96.1	17.0
Seg 5 (NS1/TuP)	29.2	56.6	35.2	49.5	28.2	33.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	0.0
Seg 6 (VP5)	13.8	36.9	13.0	31.1	17.5	37.2	97.9	99.5	99.4	99.4	99.4	99.4	99.6	99.4	99.4	0.0
Seg 7 (NS2/ViP)	0.0	29.5	0.0	19.5	28.5	0.0	93.4	96.0	74.5	99.7	100.0	99.7	99.7	99.7	99.6	35.0
Seg 8 (VP7/T13)	0.0	36.3	0.0	78.5	36.3	0.0	97.8	66.3	87.2	98.4	100.0	100.0	100.0	100.0	100.0	19.1
Seg 9 (VP6/Hel)	0.0	39.4	20.7	23.8	27.3	19.3	95.6	80.9	97.1	96.7	97.5	97.5	97.6	97.5	97.4	0.0
Seg 10 (NS3)	0.0	0.0	0.0	0.0	0.0	0.0	62.7	81.4	45.7	91.6	93.2	90.6	91.4	96.3	93.0	0.0

Table 1. Sequencing metadata and assembly information for Ross River virus (RRV) and Umatilla virus (UMAV) spiked mosquito pool samples. The reads in millions represent the number of paired, interleaved reads remaining after quality trimming. The viral reads and contigs represent all viruses in the mosquito pool sample. The number of RRV and UMAV contigs is shown, and what percentage of the virus genome is covered by these contigs. [a]RRV genome length = 11,575 bp. [b]Total UMAV genome length = 19,318 bp. UMAV segment lengths: Seg 1 = 3,711 bp; Seg 2 = 2,794 bp; Seg 3 = 2,523 bp; Seg 4 = 2,063 bp; Seg 5 = 2,107 bp; Seg 6 = 1,620 bp; Seg 7 = 1,324 bp; Seg 8 = 1,131 bp; Seg 9 = 1,104 bp; Seg 10 = 941 bp.

custom rRNA depletion, as described by manufacturer's instructions, unless where noted. The input for library preparation was 2 µL of undiluted heat-denatured RNA (total 165.2–224 ng) as preliminary experiments suggested undiluted RNA yielded more viral reads (Fig. S2A). Library preparation began with transcription of RNA into cDNA with an integrated DNase treatment. The synthesised cDNA was then sheared into 200–400 bp fragments using a S220 focused-ultrasonicator (Covaris). End repair was carried out to generate blunt ends for adaptor ligation and strand selection.

Customised insert dependent adaptor cleavage (InDA-C) ssDNA probes were used to deplete the sample of unwanted mosquito rRNA sequences. A total of 480 InDA-C probes (16–25 bp) were designed by NuGEN based on sequences provided by the authors. Specifically, these included both GenBank rRNA from a variety of mosquito species and highly abundant assembled mosquito contigs from previous metatranscriptomic sequencing of mosquito pools (a FASTA file containing the sequences used for probe design is available on Figshare: https://doi.org/10.6084/m9.figshare.9491258.v1). Preliminary experiments indicated usage of the InDA-C probes at the recommended 500 nM did not effectively deplete mosquito rRNA, however usage at 100 µM resulted in a substantial reduction of mosquito rRNA in both 100 and 1,000 mosquito pool libraries, leading to increases in virus reads (Fig. S2A,B). When used at the 100 µM concentration, the InDA-C probes were shown to reduce mosquito rRNA sequences across a range of species (Fig. S2C). Therefore, the InDA-C probes were used at a 100 µM concentration when preparing the mosquito pool samples.

After customised rRNA depletion the libraries were amplified using 14 PCR cycles and purified. All purification steps were performed using AMPure XP beads (Beckman Coulter). The size of the completed libraries was determined with a 2200 TapeStation using the D5000 ScreenTape assay (Agilent Technologies), and concentration quantified with a Qubit dsDNA HS Assay Kit. The libraries were pooled together in equimolar concentrations, diluted to 10 pM and sequenced on a HiSeq 3000 lane (Illumina) using 2 × 150 bp reads.

Analysis of metatranscriptomic sequencing data. To detect the spiked viruses in the metatranscriptomic sequencing data, reads from each individual sample were assembled into contigs using Trinity v2.4.0[34] with the read trimming (–trimmomatic) and normalisation (–normalize_reads) options selected. The assembled contigs were taxonomically classified using BLASTn v2.7.1 + with the NCBI nucleotide (nt) database (acquired 5th February 2019). BLASTn was used to identify the spiked virus contigs as it produced more specific results than BLASTx (Table S1). To determine the breadth of coverage of the spiked viruses, the assembled contigs from the individual sample reads were mapped to one set of full-length RRV and UMAV contigs using BWA-MEM v0.7.17[35] with default parameters. The BBMap pileup command[36] was used to calculate what percentage of the virus genome was covered by the contigs.

Cross-validation of the spiked virus detections was performed by mapping trimmed, interleaved reads from the individual samples to the same set of full-length RRV and UMAV contigs with BWA-MEM. Counts were derived from the alignments with the SAMtools v1.9[37] flagstat command and used to calculate reads per million (RPM). Correlation between RPM and virus spike levels was calculated using a Spearman rank correlation test with R v3.6.1[38]. The read alignments were also used to determine depth of coverage with the SAMtools depth command and visualised with the ggplot2 package v3.1.0[39] as implemented in RStudio v1.1.463[40]. The BBMap pileup command[36] was used to calculate average fold coverage of the virus genome by the reads.

The presence of other viruses in the mosquito pool was also assessed by performing a single *de novo* assembly of all the sample reads combined using Trinity. For taxonomical classification, the assembled contigs were compared to the NCBI non-redundant (nr) database (acquired 5th February 2019) using DIAMOND BLASTx v0.9.22.123[41]. BLASTx was used as opposed to BLASTn to enable detection of divergent viruses. Trimmed, interleaved reads from each individual sample were mapped to the assembled contigs from the combined sample reads with BWA-MEM and counts were summed from viral contig alignments to measure the relative abundance of viral families. Contigs were excluded from the count if they were <500 bp long, or if they also matched to the *Culex quinquefasciatus* (part of the *Culex pipiens* complex) genome (GCA_000209185.1) or the cell lines used to grow the RRV and UMAV spiked into the samples (unpublished data).

To investigate the incidence of index cross-talk among the samples, the demultiplexed reads were mapped to the PhiX genome (NC_001422.1) using BWA-MEM. PhiX is an unindexed spike-in control added to Illumina runs prior to sequencing and theoretically should not be present in the demultiplexed sample reads. Furthermore, the raw HiSeq data was re-demultiplexed using bcl2fastq Conversion Software v2.20 (Illumina) with the number of index mismatches changed from 1 to 0. The re-demultiplexed reads were also mapped to the PhiX genome.

The HiSeq FASTQ files used in this study have been uploaded to the NCBI Sequence Read Archive (SRA) under project ID PRJNA559742.

Re-sequencing of the negative control. To determine the source of contaminating RRV and UMAV reads, the negative control was re-sequenced without the spiked subsamples. First, the negative control library was re-quantified using a Qubit dsDNA HS Assay Kit. The library was then diluted to 10 pM and sequenced on a NovaSeq 6000 System (Illumina) using 2 × 150 bp reads to the same depth as the previously sequenced samples (25 million paired-end reads). The re-sequenced negative control reads were interleaved and mapped to the same set of full-length RRV and UMAV contigs as used in the analysis above.

Virus spike sample quantification using RT-ddPCR. The viral load of the 15 spiked mosquito pool samples was determined using reverse transcription droplet digital PCR (RT-ddPCR), a highly sensitive method that allows absolute quantification without the need for a standard curve[42]. The primer and probe sequences used can be found in Table S2. Double-quenched probes (Integrated DNA Technologies) were used to reduce RT-ddPCR background and increase signal intensity. The RRV primers and probe sequences were previously published[29]. The UMAV primers and probe were designed using the Primer3 algorithm in Geneious R8[43] (www.geneious.com) based on an Australian strain of UMAV using the VP2/T2 gene (NC_012755.1) reference sequence[31,33].

The One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad) was used to prepare 22 µL reaction mixtures consisting of: 5 µL of Supermix (Bio-Rad); 2 µL of reverse transcriptase (Bio-Rad); 1 µL of 300 mM dithiothreitol (DTT; Bio-Rad); 1.98 µL of each forward and reverse 10 µM virus-specific primer (Sigma-Aldrich); 0.55 µL of 10 µM virus-specific probe (Integrated DNA Technologies); 7.49 µL of UltraPure water (Invitrogen); and 2 µL of heat-denatured RNA. The reaction mixtures were loaded into an AutoDG Instrument (Bio-Rad) to generate droplets using Automated Droplet Generation Oil for Probes (Bio-Rad). The droplets were then used for RT-ddPCR using the following cycling conditions: 50 °C for 60 min; 95 °C for 10 min; 40 cycles of 95 °C for 30 s, 57 °C for 1 min; 98 °C for 10 min. After RT-ddPCR, positive and negative droplets were counted using a QX200 Droplet Reader (Bio-Rad) with FAM and HEX channels. The number of positive and negative droplets were used to calculate the concentration of RRV and UMAV as copies per µL of the final reaction (22 µL in total, including 2 µL of RNA) using QuantaSoft Software (Bio-Rad). Correlation between copies/µL and virus spike levels was calculated using a Spearman rank correlation test with R v3.6.1.

The unspiked negative control sample was also tested for RRV and UMAV using RT-ddPCR with the same specifications as above.

Virus spike sample quantification using RT-qPCR. In addition to RT-ddPCR, the viral load of the 15 spiked mosquito pool samples was measured using RT-qPCR. A one-step reaction was performed with 25 µL mixtures consisting of: 12.5 µL RT-PCR Buffer (Applied Biosystems); 1 µL of each forward and reverse 10 µM virus-specific primer (Sigma-Aldrich); 1 µL of 3.12 µM virus-specific probe (Integrated DNA Technologies); 1 µL RT-PCR Enzyme Mix (Applied Biosystems); 6 µL of UltraPure water (Invitrogen); and 2.5 µL of heat-denatured RNA. The same primer and probe sequences used for the RT-ddPCR were also used for the RT-qPCR (Table S2). The cycling conditions were as follows: 48 °C for 30 min; 95 °C for 10 min; 40 cycles of 95 °C for 15 s, 57 °C for 45 s. Correlation between Ct values and virus spike levels was calculated using a Pearson correlation test with R v3.6.1.

The unspiked negative control sample was also tested for RRV and UMAV using RT-qPCR, however instead of a probe-based assay, a SYBR-based assay was used to detect potential genetically divergent viral strains. The same reaction volumes and PCR cycle were used as above, however the 1 µL of virus-specific probe was replaced with 1 µL of 10X SYBR Green I (Invitrogen), and a melt curve protocol was added to the end of the cycle: 5 seconds at 0.5 °C increments between 65 °C and 95 °C. The negative control sample melt peak was compared to RRV and UMAV positive control melt peaks to determine if any virus was present.

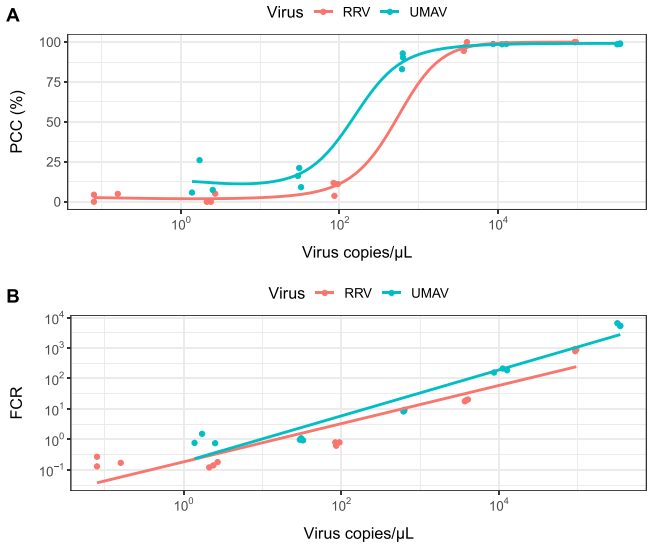


Figure 2. The relationship between copies/µL of Ross River virus (RRV) and Umatilla virus (UMAV) in the spiked mosquito samples and (**A**) percent coverage of the virus genomes by assembled contigs (PCC); (**B**) average fold coverage of the virus genomes by reads (FCR). The virus copies/µL was measured by reverse transcription droplet digital PCR (RT-ddPCR) and represents the final reaction volume (22 µL).

Results

Metatranscriptomic sequencing. A consistent level of sequence reads (mean 20.3 million per library; range 17.1–23.5 million) were obtained across the 15 spiked mosquito pool subsamples and negative control. The percentage of viral reads (mean 15.6%; range 11.6–17.1%) and number of viral contigs (mean 524; range 482–557) were also consistent across all samples (Table 1). Index cross-talk occurred during the sequencing run, with unindexed PhiX reads detectable in every sample (mean 20,944 PhiX reads, range 12,852–30,425; mean 0.05% of sample reads, range 0.03–0.08%). Re-demultiplexing the reads using more stringent parameters did not resolve the index cross-talk (Fig. S3).

Detection of the spiked viruses using metatranscriptomic sequencing was first evaluated based on the percent genome coverage of the spiked virus by assembled contigs (Percent Coverage by Contigs - PCC) (Table 1). An increase in virus input resulted in an increase in PCC for both RRV and UMAV, reaching a plateau at approximately 2.2×10^3 virus copies/µL (Fig. 2A). The 1:1 spike subsample, which was estimated to represent the RRV load of a pool of 100 mosquitoes containing a single RRV-infected mosquito, had contigs that covered the entire spiked virus genome for both RRV and UMAV. RRV 1:1 spike subsamples had a mean of 20 contigs covering a mean 100% of the genome, whereas UMAV 1:1 spike subsamples had a mean of 40 contigs covering a mean 98.9% of the genome. Contig assembly efficiency differed among the 10 UMAV segments – for example Segment 5 (NS1/TuP) assembled in every spiked sample, but Segment 10 (NS3) only assembled in the three most concentrated UMAV spike subsamples (1:400, 1:20 and 1:1).

Cross-validation of the samples was performed by mapping sample reads to the spiked virus genomes to measure average fold coverage (Fold Coverage by Reads - FCR) (Fig. 3). Like with PCC, an increase in virus input resulted in an increase in FCR for both RRV and UMAV, however FCR does not plateau like PCC does (Fig. 2B). RRV 1:1 spike subsamples had a mean 873.9 fold coverage of the genome, whereas UMAV 1:1 spike subsamples had a mean 5,778.9 fold coverage of the genome.

Sensitivity of metatranscriptomic sequencing. The PCC from the contig assembly and FCR from the read mapping approach were both used to assess the analytical sensitivity of metatranscriptomic sequencing from virus spiked mosquito pool samples. However, the determination of sensitivity was confounded by the presence of virus contigs (Table 1) and reads (Fig. 3) specific to the spiked viruses in the negative control. The contamination was unique to the metatranscriptomic sequencing as no spiked virus was detected in the negative control by RT-ddPCR or RT-qPCR. Re-sequencing of the negative control library without the spiked subsamples resulted in zero spiked virus reads, suggesting that the contamination occurred during sequencing and not during library preparation.

For the contig assembly, the three lowest RRV spike subsamples (1:400, 1:8,000 and 1:160,000) contained a mean of two contigs covering 4.6% of the RRV genome, and the negative control had 2 RRV contigs covering 3.7% of the genome. Similarly, the two lowest UMAV spike subsamples (1:8,000 and 1:160,000) contained a mean of 10.5 UMAV contigs covering 14.3% of the UMAV genome, and the negative control had six contigs covering 7.9% of the genome.

As for the read mapping approach, the two lowest RRV spike subsamples (1:8,000 and 1:160,000) had a mean 0.17 fold coverage of the genome, while the negative control had 0.11 fold coverage of the genome. For UMAV, the two lowest subsamples (1:8,000 and 1:160,000) had a mean 1.01 fold coverage of the genome, while the negative control had 0.91 fold coverage of the genome.

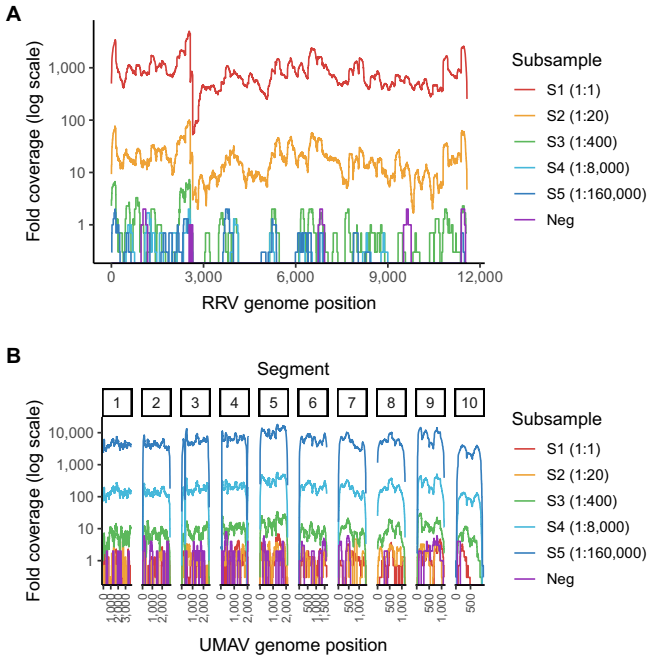


Figure 3. Mean fold coverage of the (**A**) Ross River virus (RRV) genome and (**B**) Umatilla virus (UMAV) genome across the differently spiked subsamples. RRV has a monopartite genome, whereas UMAV has a segmented genome consisting of 10 segments.

Subsample	PCC-r		FCR-r	
	RRV	UMAV	RRV	UMAV
S1 (1:1 RRV 1:160,000 UMAV)	26.7 ± 0.0*	1.7 ± 1.4	8,283.3 ± 744.0*	1.1 ± 0.5
S2 (1:20 RRV 1:8,000 UMAV)	25.9 ± 0.7*	2.0 ± 0.8*	181.1 ± 12.1*	1.1 ± 0.1
S3 (1:400 RRV 1:400 UMAV)	2.4 ± 1.2*	11.3 ± 0.7*	7.0 ± 1.0*	9.7 ± 0.6*
S4 (1:8,000 RRV 1:20 UMAV)	0.5 ± 0.8	12.6 ± 0.0*	1.4 ± 0.3	204.2 ± 30.8*
S5 (1:160,000 RRV 1:1 UMAV)	0.8 ± 0.7	12.6 ± 0.0*	1.8 ± 0.7	6,353.4 ± 731.5*

Table 2. Criteria established for detection of Ross River virus (RRV) and Umatilla virus (UMAV) in each spiked mosquito subsample. The first criterion is based on the percent genome coverage by contig ratio (PCC-r), which is calculated by dividing the percent coverage of the spiked virus genome by assembled contigs in the sample (PCC_{sample}) by the negative control (PCC_{neg}). The second criterion is based on the average fold genome coverage by reads ratio (FCR-r), which is calculated by dividing the average fold coverage of the spiked virus genome by reads in the sample (FCR_{sample}) by the negative control (FCR_{neg}). The threshold value for PCC-r and FCR-r was ≥ 2 (marked by *) and samples need both to be considered as a positive detection of either RRV or UMAV.

To address the confounding negative control results, we established a detection criterion based on PCC and FCR (Table 2). To determine if a sample was considered positive for virus spiked into the original mosquito subsample, a normalised PCC ratio (PCC-r) was calculated, where the PCC of the sample (PCC_{sample}) is divided by the negative control (PCC_{neg}). A sample with a PCC-r ≥ 2 was considered positive, which represents at least twice the level seen in the negative control. The minimum PCC_{neg} is set as 1% to prevent overinflated PCC-r values, which means $\geq 2\%$ PCC_{sample} is required for positive detection of virus. The same calculation and parameters were used for a normalised FCR ratio (FCR-r). In order for a sample to be considered positive for a virus, the PCC-r and FCR-r must both be ≥ 2.

The 1:1, 1:20, 1:400, 1:8,000 and 1:160,000 spike subsamples had an RRV PCC-r of 26.7, 25.9, 2.4, 0.5 and 0.8 respectively, and a UMAV PCC-r of 12.6, 12.6, 11.3, 2.0 and 1.7 respectively. The RRV FCR-r for the 1:1, 1:20, 1:400, 1:8,000 and 1:160,000 spike subsamples was 8,283.3, 181.1, 7.0, 1.4 and 1.8 respectively, and the UMAV FCR-r was 1.1, 1.1, 9.7, 204.2 and 6,353.4 respectively. Only the three highest spike subsamples (1:1, 1:20 and 1:400) had both PCC-r and FCR-r ≥ 2 for RRV and UMAV and were therefore considered positive for both viruses.

	Sequencing (RPM)		RT-ddPCR (copies/µL)		RT-qPCR (Ct)	
Subsample	**RRV**	**UMAV**	**RRV**	**UMAV**	**RRV**	**UMAV**
S1 (1:1 RRV 1:160,000 UMAV)	1,785.7 ± 65.4*	3.5 ± 1.2	93,766.7 ± 1,517.3	1.9 ± 0.5	17.9 ± 0.1	34.0 ± 1.1
S2 (1:20 RRV 1:8,000 UMAV)	46.8 ± 1.3*	4.0 ± 0.1	3,851.2 ± 147.8	31.3 ± 1.2	23.3 ± 0.1	30.6 ± 0.1
S3 (1:400 RRV 1:400 UMAV)	1.6 ± 0.1*	30.6 ± 2.0*	88.6 ± 4.8	625.1 ± 7.2	27.8 ± 2.0	26.2 ± 0.05
S4 (1:8,000 RRV 1:20 UMAV)	0.3 ± 0.1	619.7 ± 91.9*	2.4 ± 0.2	10,860.0 ± 1,664.8	34.1 ± 0.1	21.8 ± 0.1
S5 (1:160,000 RRV 1:1 UMAV)	0.4 ± 0.2	19,518.8 ± 281.4*	0.1 ± 0.04	336,466.7 ± 12,922.9	38.8 ± 0.9	16.6 ± 0.1
Negative control	0.1	1.4	0	0	0	0

Table 3. Comparison of Ross River virus (RRV) and Umatilla virus (UMAV) quantification in the spiked mosquito subsamples and negative control using metatranscriptomic sequencing, reverse transcription droplet digital PCR (RT-ddPCR) and quantitative PCR (RT-qPCR). The sequencing results are shown as mapped reads per million (RPM), with the subsamples considered positive marked by an asterisk (based on having percent coverage by contig ratio (PCC-r) and average fold coverage by reads ratio (FCR-r) both ≥ 2). The RT-ddPCR measurement refers to copies per µL of the final reaction (22 µL in total). Aside from the negative control, all results are shown as mean with one standard deviation based on three technical replicates.

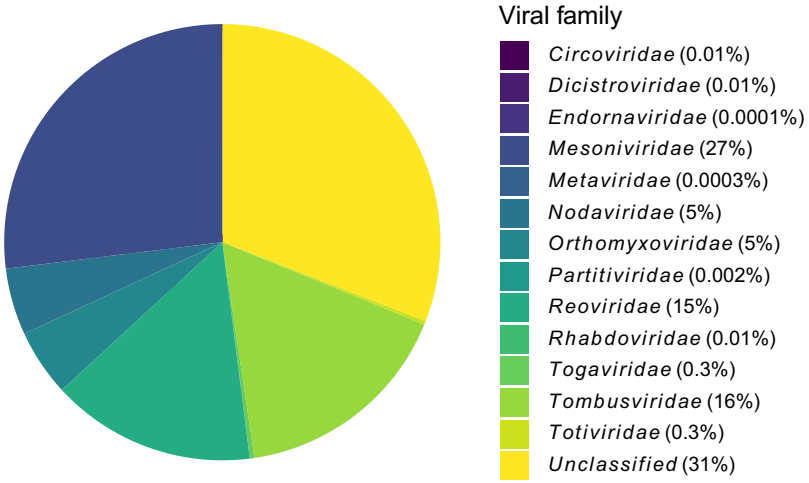


Figure 4. The viral family composition of the pool of 100 *Culex australicus* mosquitoes, shown as percentage of total read counts for each family.

Specificity of metatranscriptomic sequencing. Specificity was based on the accuracy of taxonomic classification of the spiked virus contigs assembled for each sample. The BLASTn search of the nt database correctly identified all RRV contigs from the individual samples as RRV, whereas all but two (0.4%) of the UMAV contigs were correctly identified as UMAV (Table S1B). Those two contigs were identified as Koyama Hill virus (KHV), which is also a member of the *Umatilla virus* species[44]. On closer inspection of the two contigs it was found the sequences matching KHV were in an untranslated region (UTR), and the UMAV reference in the nt database did not contain UTR sequences.

Comparison of virus detection methods. To compare the three virus detection methods: RPM from the metatranscriptomic sequencing results; the copies/µL measurement from the RT-ddPCR; and the cycle threshold (Ct) value from the RT-qPCR were used (Table 3). Virus spike levels positively correlated with RPM ($R = 0.927$, $p = < 0.001$) and copies/µL ($R = 0.982$, $p = < 0.001$), and negatively correlated with Ct ($R = -0.76$, $p = 0.002$). The lowest concentration spike subsample (1:160,000) was detectable by both RT-ddPCR and RT-qPCR, with mean 0.1 copies/µL and Ct 38.8 for RRV, and mean 1.9 copies/µL and Ct 34.0 for UMAV. Based on the PCC-r and FCR-r criterion, only the three highest spike subsamples were considered positive for RRV and UMAV (1:1, 1:20 and 1:400). The lowest of these (1:400) corresponded to mean 1.6 RPM, 88.6 copies/µL and Ct 27.8 for RRV, and mean 30.6 RPM, 625.1 copies/µL and Ct 26.2 for UMAV.

Detection of other viruses. In addition to the two spiked viruses, metatranscriptomic sequencing revealed the presence of other viruses in the pool of 100 *Cx. australicus* mosquitoes (Fig. 4). The most abundant assembled virus contigs were classified as *Mesoniviridae* (27%), *Tombusviridae* (16%) and *Reoviridae* (15%) or were unclassified (31%). Nineteen previously characterised viruses were present in the pool (Table S3) all of which have been detected in mosquito samples and are currently considered to be insect-specific.

Discussion

In this study, we used a spiking experiment to investigate the analytical sensitivity and specificity of a metatranscriptomic pipeline in detecting two RNA viruses with differing genome structure in a pool of 100 *Cx. australicus* mosquitoes (Fig. 1). The metatranscriptomic pipeline successfully sequenced the full-length genome of both viruses in the spiked subsample that corresponded to a biologically relevant viral load representing a single RRV-infected mosquito in a pool of 100 mosquitoes (1:1 spike dilution). Detection of RRV in a pool of 1,000 mosquitoes containing one RRV-infected mosquito was also demonstrated (Fig. S2B). This level of sensitivity was achieved by using a customised mosquito rRNA depletion, which helped to achieve a higher portion of viral sequencing reads (11.7–17.3%) compared to other metatranscriptomic studies where mosquito pool samples had as little as <1% viral reads[45–47]. Notably, the rRNA depletion was effective only when a higher concentration of the customised probe mixture was used than advised in the manufacturer's instructions (Fig. S2A,B). It is possible this is due to the particularly high level of host RNA when using a large pool of mosquitoes as starting material. Other approaches that helped to increase sensitivity were the use of undiluted RNA as input for library preparation (Fig. S2A) and RNA heat-denaturation to improve detection of dsRNA genomes (data not shown). All these approaches are recommended to improve sensitivity when performing metatranscriptomic sequencing of mosquito traps for arbovirus surveillance.

In order to increase accuracy, detection of the spiked virus genomes in the metatranscriptomic data was first performed using a contig assembly approach from which Percent Coverage by Contigs (PCC) was derived, and then cross-validated with read mapping to the virus genomes, from which Fold Coverage by Reads (FCR) was derived. While there was a strong relationship between PCC/FCR and the viral copies/μL (Fig. 2A,B), the presence of contigs and reads specific to the spiked viruses in the negative control confounded detection at lower viral loads (Table 1 and Fig. 3). Re-sequencing of the negative control library returned zero spiked virus reads, indicating that the contamination occurred during sequencing and not during library preparation. It is possible the contaminating reads are a result of index cross-talk, which occurs when reads are misassigned due to incorrect matching of the indexes used to multiplex samples[48]. The presence of PhiX in the sample reads also indicate index cross-talk occurred, since the PhiX spike-in is unindexed and therefore should not be present in any of the demultiplexed samples. Index cross-talk can be caused by spreading of signal on flow cells, sequencing errors introduced during bridge amplification, improper cluster resolution and misread indexes[48]. The rate of index cross-talk increases with the use of Illumina patterned flow cells, and also single indexes[49], both of which were employed in this study. Using unique dual indexes to multiplex samples has been shown to significantly reduce index cross-talk, thereby increasing the sensitivity of sequencing[48,49]. Therefore, the use of both negative controls and unique dual indexes is recommended when using metatranscriptomics for sensitive applications such as surveillance in order to improve detection and dependability of the results.

To account for the contaminating reads in the negative control, we established a detection criterion where the PCC_{sample} and FCR_{sample} for a virus is divided by the PCC_{neg} and FCR_{neg}, respectively, to produce normalised ratios (PCC-r and FCR-r). Both ratios must be ≥ 2 for that sample to be considered positive for a virus. Using this criterion, only the three highest spike subsamples (1:1, 1:20 and 1:400) were positive for both RRV and UMAV (Table 2). The maximum FCR-r value for RRV was higher than for UMAV (8,283.3 vs. 6,353.4), which was due to the negative control containing less RRV reads than UMAV reads (FCR_{neg} 0.11 vs. 0.91). This pattern was also present in the RRV and UMAV PCC-r and PCC_{neg} values. The higher UMAV contamination in the negative control was due to the subsamples having higher concentrations of UMAV than RRV (Table 3), leading to increased index cross-talk[50]. The higher UMAV concentrations also meant that the PCC plateaued earlier for UMAV than for RRV (Fig. 2A). The PCC-r and FCR-r values give an indication of the virus genome assembly and coverage, and virus concentration present in a sample, respectively, while accounting for contamination in the negative control. Patterns in these values can be useful for surveillance, for instance, a high PCC and low FCR suggests a virus is present but at low concentration. Conversely, a low PCC and high FCR could be indicative of a related virus or erroneous reference genome[51]. This approach is dependent on complete genome sequences in the reference database for accuracy, and care needs to be taken when analysing segmented viruses to ensure coverage is calculated for the whole genome and not just one segment. As with any detection tool, it is recommended that any viruses of public health concern detected by metatranscriptomic sequencing are confirmed using alternative virus detection methods such as PCR[52].

Further studies utilising simulated samples with a finer scale of virus concentration and more negative samples will allow validation of the cut-off values for the PCC-r and FCR-r criterion using a robust statistical-based approach. The proposed value of ≥ 2 means positive detections are at least twice the level seen in the negative control, which has been previously used in other diagnostic tests such as PCR[53] and ELISA[54], however remains arbitrary when based on only one negative control sample. Future metatranscriptomic sequencing of mosquito pools that are known to be positive for arboviruses can also be used for further evaluation of the criterion and will improve this approach as a routine surveillance tool.

When investigating the assembly of the 10 UMAV genome segments, we found that certain segments assembled in every sample, while other segments, regardless of segment size, only assembled in higher concentration spike samples (Table 1). When reads were mapped to the UMAV genome all 10 segments had similar coverage for each spiked subsample (Fig. 3B), suggesting the difference in assembly efficiency does not have to do with the availability of the segments in the samples but rather with the contig assembly analysis. Often the inability to detect all of a segmented virus genome suggests the segments are highly divergent from previously sequenced viruses[55], however the reference genome for the UMAV strain spiked into the mosquito subsamples was in the NCBI nt database used for analysis. Segment 5 (NS1/TuP) was the most frequently assembled segment (29.2–100% PCC across all samples) and interestingly, has the longest UTR sequence that has been recorded for an orbivirus[31]. It is possible the varying lengths of the UTR sequences for each segment may have affected the assembly[56].

Further work to investigate this anomaly could include the comparison of different transcriptome assemblers when working with segmented genomes.

A consistently high specificity was achieved by assembling and taxonomically identifying the spiked viruses, with 100% of RRV and 99.6% of UMAV contigs identified correctly. The misidentification of two UMAV contigs as KHV (also a member of the *Umatilla virus* species) occurred because the UMAV genome in the nt database did not contain any UTR sequences, demonstrating how incomplete reference databases can lead to errors. The specificity was also dependent on the taxonomic classification approach – when BLASTx (translated nucleotide query to protein database) was used instead of BLASTn (nucleotide query to nucleotide database), it led to a decrease in specificity (Table S1B). BLASTx is commonly used in metatranscriptomic data analysis as it can detect divergent sequences which enables novel virus discovery, however BLASTn produces less erroneous results and therefore may be more suited for known pathogen identification[57]. Often studies take a combined approach that utilises both nucleotide and protein information to achieve more accurate and sensitive virus classification[58–60].

Metatranscriptomic sequencing was not as sensitive as RT-ddPCR and RT-qPCR in detecting the spiked viruses, with both PCR methods successfully detecting RRV and UMAV in all of the spiked mosquito pool subsamples without producing background noise in the negative control (Table 3). Eliminating the contaminant sequences in the negative control would improve the sensitivity of metatranscriptomic sequencing. However, other factors can also affect sensitivity, including the type of sample being used. Metatranscriptomic sequencing has reached a virus detection limit similar to diagnostic qPCR when liquid biological samples are being used, such as blood[61,62], nasopharyngeal swabs[63,64] or clarified cell culture supernatant[65]. However, when complex samples such as sewage[1,66] or plant tissue[67] are used, metatranscriptomic sequencing is considerably less sensitive. Despite this reduced sensitivity, it is important to note metatranscriptomic sequencing can detect multiple regions, if not the entire virus genome (Fig. 3), whereas PCR targets only a small region. Acquiring more genomic information enables detection of viruses that may evade PCR due to sequence divergence in the diagnostic region and can also be used for molecular epidemiology to gain insight into viral emergence and spread during an outbreak. The utility of this approach was recently evidenced in Nigeria during a Lassa fever outbreak, where metatranscriptomic sequencing on a MinION sequencer enabled simultaneous detection and characterisation of Lassa virus, a highly variable RNA virus that poses difficulties for PCR-based diagnostics[4]. The use of whole genome information is highly beneficial for surveillance not only to describe the diversity of viruses circulating, but also to understand where they came from, how they will be transmitted, and how different strains have evolved over time.

The nontargeted nature of metatranscriptomics meant that not only were the whole genomes of the spiked viruses sequenced, so were other viruses present in the pool of 100 *Cx. australicus* mosquitoes. *De novo* assembly revealed a variety of viral families (Fig. 4), which included 19 previously characterised viruses (Table S3). These results are consistent with prior metatranscriptomic studies, with 15 of the viruses identified in Australian mosquitoes, and 11 of those from the Shi *et al*. study[11]. This is the first time Culex circovirus-like virus, Culex Hubei-like virus, Culex-associated Tombus-like virus and Yongsan picorna-like virus 2 have been detected in Australia. The detection of a circovirus (ssDNA virus) confirms that the metatranscriptomic protocol used is capable of sequencing DNA viruses, despite being targeted at RNA viruses. With DNase-treated RNA as the input material it is possible this is mRNA produced by the circovirus, and it could also be DNA if the DNase treatment was not 100% efficient[68]. A recent study on contaminating viral sequences in virome data suggests circovirus-like viruses are a common contaminant derived from laboratory components[69]. Other types of DNA viruses would need to be tested to determine if this protocol can detect both RNA and DNA without separate nucleic acid library preparations. Whilst the known viruses identified in this pool of mosquitoes are not known to cause disease in mammalian cells, the ability to detect these viruses without targeting them highlights the value of metatranscriptomic sequencing in arbovirus surveillance.

The wealth of information provided by metatranscriptomic sequencing enhances arbovirus surveillance, however this tool needs to be affordable in order to be broadly utilised in surveillance programs. Processing a sample with the same commercial kits and depth of sequencing used in this study costs approximately AUD$230. Over half of this cost is attributed to the library preparation with customised rRNA depletion, and could be reduced by using a cheaper kit (e.g. NEBNext Ultra II RNA) and an in-house depletion method, such as the Cas9-based approach described in Gu *et al*.[70]. The second largest cost is sequencing, with the ~20 million reads per sample used in this study costing approximately AUD$100 using an Illumina NovaSeq sequencer[71]. This depth of sequencing enabled detection of RRV in the 1:400 spike subsample, which is equivalent to 1 positive mosquito in 40,000, therefore the sequencing depth and cost could be halved whilst remaining considerably sensitive. These suggested changes lower the overall cost per sample to approximately AUD$110. This cost does not include labour time, which amounts to approximately three days for the nucleic acid extraction and library preparation of 32 samples. Automation of some of the steps could increase the number of samples processed simultaneously. The NovaSeq run time is 40 hours[71], resulting in a week turnaround time. Due to the cost and time involved, metatranscriptomic sequencing is currently suited as an additional tool to routine surveillance, providing in-depth information on viral activity in mosquito populations at regular intervals throughout the season, perhaps on a monthly basis. It is likely the time and cost associated with metatranscriptomic sequencing will decrease in the future, allowing it to be used more routinely.

This study has provided information on the sensitivity and specificity of metatranscriptomic sequencing for detection of arboviruses in large pools of mosquitoes, which is essential for the incorporation of this technique into arbovirus surveillance programs. Metatranscriptomic sequencing successfully detected a virus in a pool of 100 mosquitoes at biologically relevant levels, and also in a pool of 1,000 mosquitoes (Fig. S2B). While metatranscriptomic sequencing was less sensitive than diagnostic gold standard approaches such as RT-qPCR and RT-ddPCR, it provided more in-depth information by spanning the entire virus genome, and detecting all viruses present in the mosquito pool. Choices made during the laboratory process and bioinformatic analysis affected the sensitivity and specificity of virus detection, and therefore standardised protocols for both processes need to

be established for routine use of metatranscriptomic sequencing. The criterion for positive detection of a virus established in this paper is one example of a process that can be applied to produce comparable results, which also accounts for potential contamination found in the negative control. Further work utilising wild caught mosquitoes from diverse populations will help to establish metatranscriptomic sequencing as a tool that can broaden the capabilities of arbovirus surveillance.

Data availability

The sequences used for the customised mosquito rRNA probe design are available as a FASTA file on Figshare: https://doi.org/10.6084/m9.figshare.9491258.v1. The unprocessed FASTQ files from the Illumina HiSeq are available on the NCBI SRA Database under project ID PRJNA559742.

Received: 23 August 2019; Accepted: 29 November 2019;
Published online: 18 December 2019

References

1. Fernandez-Cassi, X. *et al*. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Sci. Total Environ.* **618**, 870–880 (2018).
2. Harvey, E. *et al*. Extensive diversity of RNA viruses in Australian ticks. *J. Virol.* **93**, e01358–18 (2019).
3. Zheng, X. *et al*. Viral metagenomics of six bat species in close contact with humans in southern China. *Arch. Virol.* **163**, 73–88 (2018).
4. Kafetzopoulou, L. E. *et al*. Metagenomic sequencing at the epicenter of the Nigeria 2018 Lassa fever outbreak. *Science* **363**, 74–77 (2019).
5. Bhatt, S. *et al*. The global distribution and burden of dengue. *Nature* **496**, 504–507 (2013).
6. Gossner, C. M. *et al*. West Nile virus surveillance in Europe: moving towards an integrated animal-human-vector approach. *Euro Surveill.* **22**, (2017).
7. Knope, K. *et al*. Arboviral diseases and malaria in Australia, 2014–15: Annual report of the National Arbovirus and Malaria Advisory Committee. *Commun. Dis. Intell. (2018)* **43**, (2019).
8. Oliver, J. *et al*. Twenty years of surveillance for Eastern equine encephalitis virus in mosquitoes in New York State from 1993 to 2012. *Parasit. Vectors* **11**, 362 (2018).
9. Greninger, A. L. A decade of RNA virus metagenomics is (not) enough. *Virus Res.* **244**, 218–229 (2018).
10. Chandler, J. A., Liu, R. M. & Bennett, S. N. RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Front. Microbiol.* **6**, 185 (2015).
11. Shi, M. *et al*. High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J. Virol.* **91**, e00680–17 (2017).
12. Xiao, P. *et al*. Metagenomic sequencing from mosquitoes in China reveals a variety of insect and human viruses. *Front. Cell. Infect. Microbiol.* **8**, (2018).
13. Bigot, D. *et al*. Discovery of *Culex pipiens* associated Tunisia virus: a new ssRNA(+) virus representing a new insect associated virus family. *Virus Evol.* **4**, vex040 (2018).
14. Hall-Mendelin, S., Allcock, R., Kresoje, N., van den Hurk, A. F. & Warrilow, D. Detection of arboviruses and other micro-organisms in experimentally infected mosquitoes using massively parallel sequencing. *PLoS One* **8** (2013).
15. Batovska, J., Lynch, S. E., Rodoni, B. C., Sawbridge, T. I. & Cogan, N. O. Metagenomic arbovirus detection using MinION nanopore sequencing. *J. Virol. Methods* **249**, 79–84 (2017).
16. Sadeghi, M. *et al*. Virome of >12 thousand Culex mosquitoes from throughout California. *Virology* **523**, 74–88 (2018).
17. Ribeiro, G. *et al*. Detection of RNA-dependent RNA polymerase of Hubei reo-like virus 7 by next-generation sequencing in *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes from Brazil. *Viruses* **11** (2019).
18. Conceição-Neto, N. *et al*. Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Sci. Rep.* **5**, 16532 (2015).
19. Rosseel, T., Ozhelvaci, O., Freimanis, G. & Van Borm, S. Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. *J. Virol. Methods* **222**, 72–80 (2015).
20. Fauver, J. R. *et al*. A reverse-transcription/RNase H based protocol for depletion of mosquito ribosomal RNA facilitates viral intrahost evolution analysis, transcriptomics and pathogen discovery. *Virology* **528**, 181–197 (2019).
21. Kukutla, P., Steritz, M. & Xu, J. Depletion of ribosomal RNA for mosquito gut metagenomic RNA-seq. *J. Vis. Exp*. doi:10.3791/50093 (2013).
22. Schlaberg, R. *et al*. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch. Pathol. Lab. Med.* **141**, 776–786 (2017).
23. McHardy, A. C., Martín, H. G., Tsirigos, A., Hugenholtz, P. & Rigoutsos, I. Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Methods* **4**, 63–72 (2007).
24. Miller, S. *et al*. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res.* **29**, 831–842 (2019).
25. Rohe, D. L. & Fall, R. P. A miniature battery powered CO2 baited light trap for mosquito borne encephalitis surveillance. *Bull. Soc. Vector Ecol.* **4**, 24–27 (1979).
26. Dobrotworsky, N. V. *The Mosquitoes of Victoria*. Melbourne University Press, (1965).
27. Russell, R. C. *A Colour Photo Atlas of Mosquitoes of Southeastern Australia*. (The Department of Medical Entomology, Westmead Hospital and the University of Sydney, 1996).
28. Doherty, R. L. Ross River virus. in *Catalogue of Arthropod-borne Viruses of the World* vol. 1760:421 Department of Health, Education, and Welfare, (1967).
29. Hall, R., Prow, N. & Pyke, A. Ross River Virus. in *Molecular Detection of Human Viral Pathogens* 349–359 CRC Press, (2011).
30. Lynch, S. E. *et al. Victorian Arbovirus Disease Control Program Annual Report 2015–2016*. Agriculture Victoria, (2016).
31. Belaganahalli, M. N. *et al*. Umatilla virus genome sequencing and phylogenetic analysis: Identification of Stretch Lagoon orbivirus as a new member of the Umatilla virus species. *PLoS One* **6**, e23605 (2011).
32. Batovska, J. *et al*. Effective mosquito and arbovirus surveillance using metabarcoding. *Mol. Ecol. Resour.* **18**, 32–40 (2018).
33. Cowled, C. *et al*. Genetic and epidemiological characterization of Stretch Lagoon orbivirus, a novel orbivirus isolated from *Culex* and *Aedes* mosquitoes in northern Australia. *J. Gen. Virol.* **90**, 1433–1439 (2009).
34. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
35. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]* (2013).
36. Bushnell, B. *BBMap short read aligner, and other bioinformatic tools*. (2017).
37. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
38. R Core Team. *R: A language and environment for statistical computing*. (2019).

39. Wickham, H. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, (2009).
40. RStudio Team. *RStudio: Integrated development for R*. (2015).
41. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
42. Abachin, E. *et al*. Comparison of reverse-transcriptase qPCR and droplet digital PCR for the quantification of dengue virus nucleic acid. *Biologicals* **52**, 49–54 (2018).
43. Kearse, M. *et al*. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
44. Ejiri, H. *et al*. First isolation and characterization of a mosquito-borne orbivirus belonging to the species Umatilla virus in East Asia. *Arch. Virol.* **159**, 2675–2685 (2014).
45. Cholleti, H. *et al*. Discovery of novel viruses in mosquitoes from the Zambezi Valley of Mozambique. *PLoS One* **11**, e0162751 (2016).
46. Fauver, J. R. *et al*. West African *Anopheles gambiae* mosquitoes harbor a taxonomically diverse virome including new insect-specific flaviviruses, mononegaviruses, and totiviruses. *Virology* **498**, 288–299 (2016).
47. Xiao, P. *et al*. Metagenomic analysis of *Flaviviridae* in mosquito viromes isolated from Yunnan Province in China reveals genes from dengue and Zika viruses. *Front. Cell. Infect. Microbiol.* **8** (2018).
48. MacConaill, L. E. *et al*. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* **19**, 30 (2018).
49. Costello, M. *et al*. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genom.* **19** (2018).
50. Larsson, A. J. M., Stanley, G., Sinha, R., Weissman, I. L. & Sandberg, R. Computational correction of index switching in multiplexed sequencing libraries. *Nat. Methods* **15**, 305–307 (2018).
51. Pickrell, J. K., Gaffney, D. J., Gilad, Y. & Pritchard, J. K. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics* **27**, 2144–2146 (2011).
52. Liu, S., Vijayendran, D. & Bonning, B. C. Next generation sequencing technologies for insect virus discovery. *Viruses* **3**, 1849–1869 (2011).
53. Courtney, B. C., Smith, M. M. & Henchal, E. A. Development of internal controls for probe-based nucleic acid diagnostic assays. *Anal. Biochem.* **270**, 249–256 (1999).
54. Mboloi, M. M., Bekker, C. P. J., Kruitwagen, C., Greiner, M. & Jongejan, F. Validation of the indirect MAP1-B enzyme-linked immunosorbent assay for diagnosis of experimental Cowdria ruminantium infection in small ruminants. *Clin. Diagn. Lab. Immunol.* **6**, 66–72 (1999).
55. Cook, S. *et al*. Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* **8**, e80720 (2013).
56. Zhang, R. *et al*. A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* **45**, 5061–5073 (2017).
57. Bibby, K., Viau, E. & Peccia, J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Lett. Appl. Microbiol.* **52**, 386–392 (2011).
58. Ajami, N. J., Wong, M. C., Ross, M. C., Lloyd, R. E. & Petrosino, J. F. Maximal viral information recovery from sequence data using VirMAP. *Nat. Commun.* **9**, 3205 (2018).
59. Carissimo, G. *et al*. Identification and characterization of two novel RNA viruses from *Anopheles gambiae* species complex mosquitoes. *PLoS One* **11**, e0153881 (2016).
60. Ho, T. & Tzanetakis, I. E. Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* **471–473**, 54–60 (2014).
61. Cheval, J. *et al*. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J. Clin. Microbiol.* **49**, 3268–3275 (2011).
62. Frey, K. G. *et al*. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genom.* **15**, 96 (2014).
63. Graf, E. H. *et al*. Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: A systematic comparison to a commercial PCR panel. *J. Clin. Microbiol.* **54**, 1000–1007 (2016).
64. Greninger, A. L. *et al*. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLOS ONE* **5**, e13381 (2010).
65. Li, L. *et al*. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J. Virol. Methods* **213**, 139–146 (2015).
66. Bibby, K. & Peccia, J. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environ. Sci. Technol.* **47**, 1945–1951 (2013).
67. Fernandez-Cassi, X. *et al*. A metagenomic assessment of viral contamination on fresh parsley plants irrigated with fecally tainted river water. *Int. J. Food Microbiol.* **257**, 80–90 (2017).
68. Hall, R. J. *et al*. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods* **195**, 194–204 (2014).
69. Asplund, M. *et al*. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clin. Microbiol. Infect.* **25**, 1277–1285 (2019).
70. Gu, W. *et al*. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* **17**, 41 (2016).
71. Piper, A. M. *et al*. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *Gigascience* **8** (2019).

Acknowledgements

The authors thank Russell Barnes for collecting the mosquitoes and Karen Brown for morphologically identifying them; Jonathan Darbro and Elise Kho for providing the Ross River virus-infected *Aedes notoscriptus* specimens; and Tracie Webster for her assistance in sequencing the samples and re-demultiplexing the reads. The authors also thank the two anonymous reviewers whose important feedback helped to improve this manuscript. This work was funded by the Biosciences Research Innovation Fund Program provided by the Victorian Department of Jobs, Precincts and Regions. The mosquito trapping was performed as part of the Victorian Arbovirus Disease Control Program, which is funded by the Department of Health and Human Services. JB is supported by an Australian Government Research Training Program Scholarship.

Author contributions

All authors contributed to the conception and design of the experiment; P.T.M. and J.B. prepared the viral spikes; J.B. performed the nucleic acid extractions, prepared the sequencing libraries, analysed the data, and wrote the manuscript; all authors contributed to the editing of the final manuscript and approved the version submitted for publication.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-55741-3.

Correspondence and requests for materials should be addressed to J.B. or S.E.L.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

Supplementary Information

Estimating virus isolate concentration using RT-ddPCR

The virus isolates that were spiked into the mosquito homogenate subsamples were first quantified using reverse transcription droplet digital PCR (RT-ddPCR). The method and calculations used are described below using the Ross River virus (RRV) isolate as an example.

Lab method

RNA was extracted from the RRV isolate using a QIAamp Viral RNA Mini Kit and quantified using an RNA HS assay on the Qubit:

- Qubit reading = 4.58 ng/µL

The RNA was then diluted 1:1,000 for input into RT-ddPCR, which was performed using a One-Step RT-ddPCR Advanced Kit for Probes. The diluted RNA input was 2 µL and the sample mix was 22 µL in total.

- RT-ddPCR reading = 69.035 copies/µL

In order to normalise it with the Umatilla virus (UMAV) isolate, the RRV isolate was diluted by a factor of 2.4 and spiked into the S1 mosquito homogenate subsample. The calculation used to estimate the concentration of the spiked RRV isolate is shown below.

Calculation

Start with RT-ddPCR reading:

- 69.035 copies/µL

Number of copies of cDNA/RNA in the entire 22 µL reaction:

- 69.035 x 22 = 1,518.774 copies

- Note: The assumption here is that conversion efficiency from RNA to cDNA is 100%.

Concentration of input RNA (2 µL used):

- 1,518.774 / 2 = 759.387 copies/µL

Concentration of RNA before 1:1,000 dilution:

- 759.387 x 1,000 = 759,387 copies/µL

Calculate copies/ng of RNA:

- Qubit reading = 4.58 ng/µL
- 759,387 / 4.58 = 165,805 copies/ng of RNA
- Note: The assumption here is that all the RNA measured by the Qubit belongs to RRV.

Adjust for final dilution:

- 165,805 / 2.4 = 68,952.88 copies/ng of RNA

  $= 6.9 \times 10^4$ copies/ng of RNA

Table S1: A comparison of the number of assembled contigs from all individual samples classified as **A)** '*Togaviridae*' or **B)** '*Orbivirus*' using a BLASTn search of the NCBI nucleotide database, and a BLASTx search of the non-redundant database. Closely related viruses are grouped together by different colours.

A)

BLASTn results		BLASTx results	
No. of contigs	**Taxonomic ID**	**No. of contigs**	**Taxonomic ID**
131	Ross River virus	123	Ross River virus
		1	Getah virus

B)

BLASTn results		BLASTx results	
No. of contigs	**Taxonomic ID**	**No. of contigs**	**Taxonomic ID**
339	Anopheles hinesorum orbivirus	369	Anopheles hinesorum orbivirus
246	Anopheles annulipes orbivirus	236	Anopheles annulipes orbivirus
449	Umatilla virus M4941_15	54	Morris orbivirus
2	Koyama Hill virus	339	Umatilla virus M4941_15
		51	Stretch Lagoon orbivirus
		48	Koyama Hill virus
		4	Umatilla virus USA1969/01
		1	Minnal virus

Table S2: Primer and probe sequences used for reverse transcription qPCR and ddPCR quantification of Ross River virus (RRV) and Umatilla virus (UMAV) in clarified mosquito subsamples.

Name	Sequence (5'–3')	Amplicon size (bp)	Gene	Reference
RRVE2F	ACGGAAGAAGGGATTGAGTACCA	67	E2	(Hall et al., 2011)
RRVE2R	TCGTCAGTTGCGCCCATA		E2	(Hall et al., 2011)
RRV_E2_ZEN_IB	56-FAM/CAACAACCC/ZEN/GCCGGTCCGC/3IABkFQ	N/A	E2	(Hall et al., 2011)
UMAV_84_F	CAGAGAGATGACTATCGACG	84	Seg 2 (VP2/T2)	This study
UMAV_84_R	TTGTAGGTTCCGATCATAGG		Seg 2 (VP2/T2)	This study
UMAV_84_ZEN_IB	56-FAM/CACAAGCAT/ZEN/GGTTACGTACATATTC/3IABkFQ	N/A	Seg 2 (VP2/T2)	This study

Table S3: Known viruses present in the pool of 100 *Culex australicus* mosquitoes as identified by de novo assembly. The length of the assembled contigs is shown along with the amino acid percentage identity (AA PI%) to the reference sequence.

Viral family	Virus	Contig length/s (bp)	AA PI%
Circoviridae	Culex circovirus-like virus†	1,256-1,472	96.9-97.2
Mesoniviridae	Ngewotan virus*	20,244	99.4
Nodaviridae	Culex Hubei-like virus†	668	97.6
Orthomyxoviridae	Wuhan Mosquito Virus 6*	921-1.554	99.7-100
Reoviridae	Anopheles annulipes orbivirus	4,414	99.6
Rhabdoviridae	Beaumont virus	4,395	96.8
Tombusviridae	Culex-associated Tombus-like virus†	1,752	99.6
Totiviridae	Australian Anopheles totivirus	923-6,142	67.9-97.3
Unclassified	Castlerea virus	9,456	100
Unclassified	Culex mononega-like virus 1*	1,681-5,974	99.8-100
Unclassified	Culex mononega-like virus 2*	8,668-13,297	99.8-100
Unclassified	Culex negev-like virus 1*	11,002	99.6
Unclassified	Culex negev-like virus 3*	1,494-7,299	99.6-100
Unclassified	Culex phasma-like virus*	1,886-5,205	100
Unclassified	Culex rhabdo-like virus*	11,455	100
Unclassified	Hubei chryso-like virus 1*	3,149-3,370	99.8-100
Unclassified	Hubei reo-like virus 7*	3,717	96.3
Unclassified	Yongsan picorna-like virus 2†	9,781	96.4
Unclassified	Zhejiang mosquito virus 3*	2,834	96.7

***** Viruses previously detected in Western Australia mosquitoes (Shi et al., 2017).

† Viruses detected for the first time in Australia.

Fig. S1: The reverse transcription quantitative PCR (RT-qPCR) amplification curves for Ross River virus (RRV) in: **A)** A pool of 100 *Culex australicus* mosquitoes containing a single RRV-infected mosquito (run in triplicate); **B)** The RRV-spiked mosquito subsample (S1 – S5) replicates (run in duplicate). The S1 subsample (1:1 spike dilution) is representative of the viral load found in A) and is marked on the amplification plot.

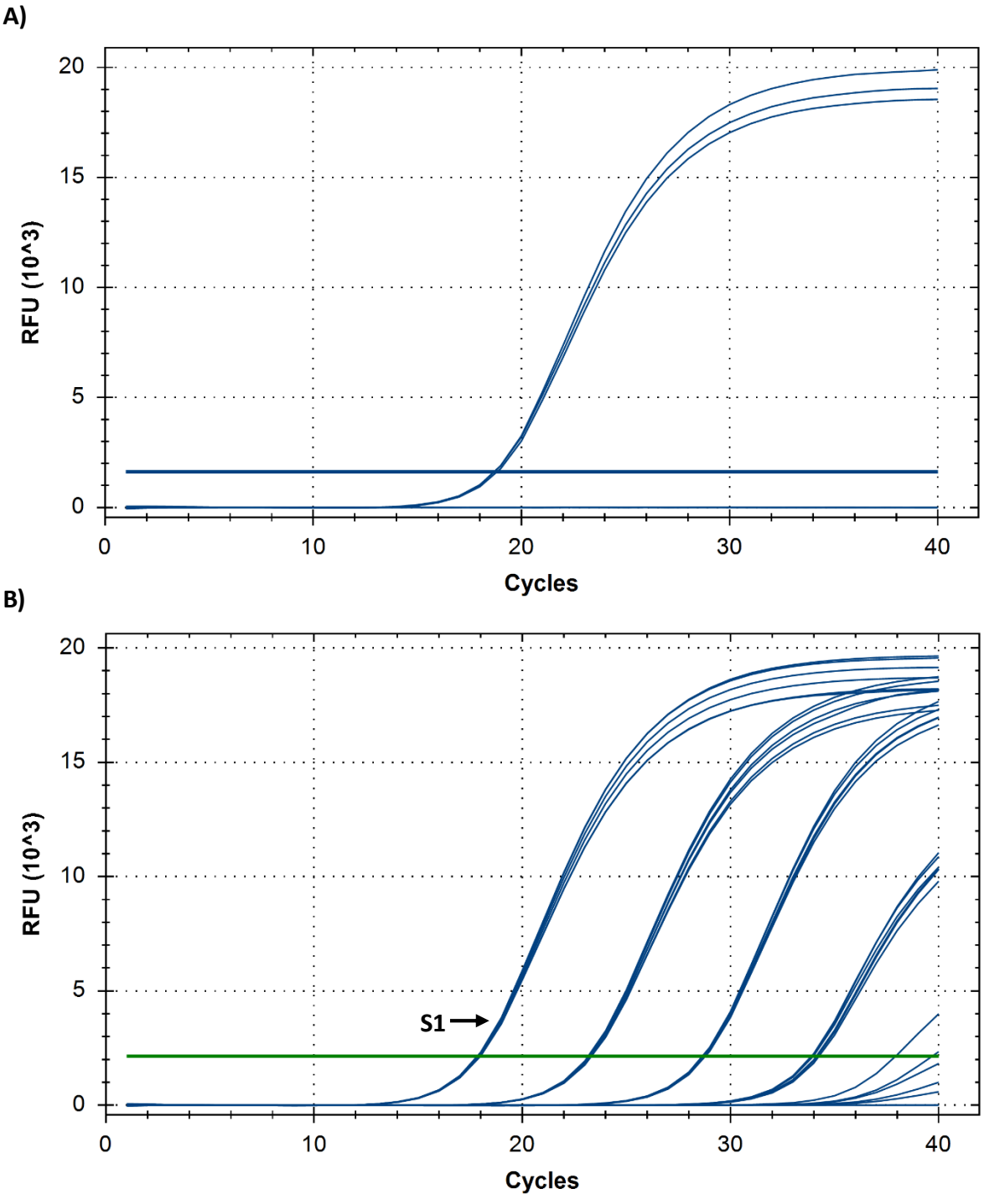
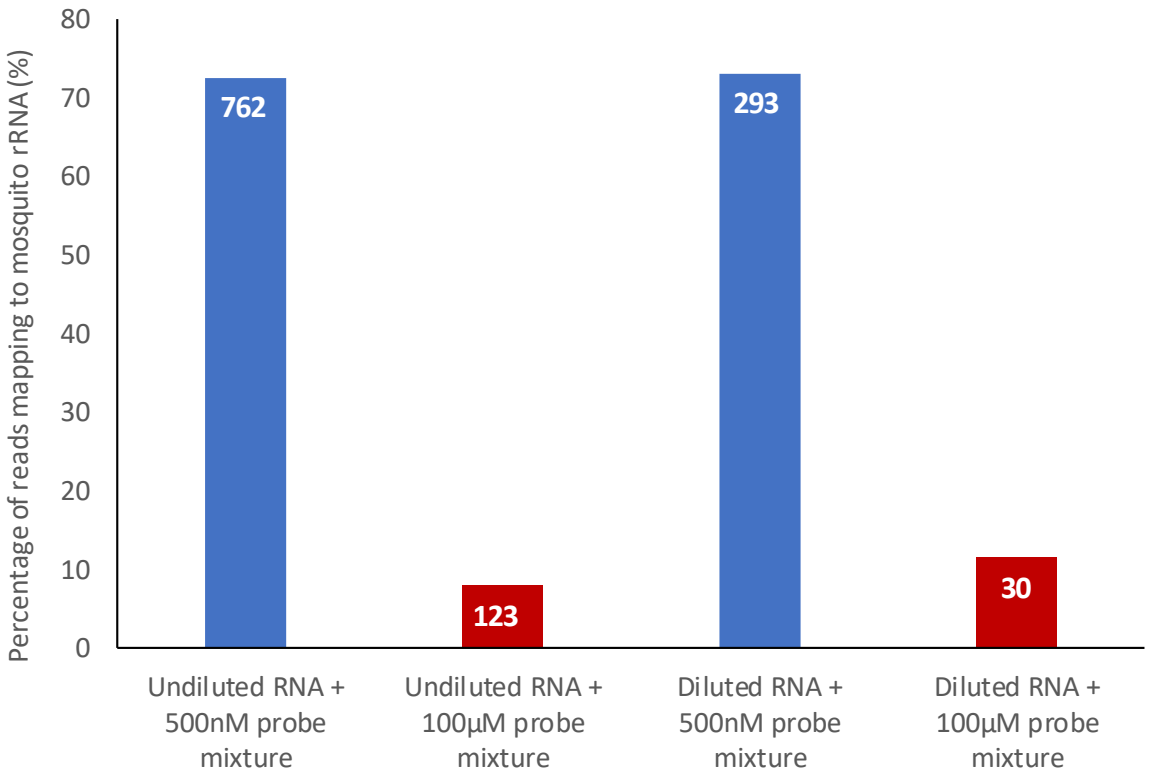
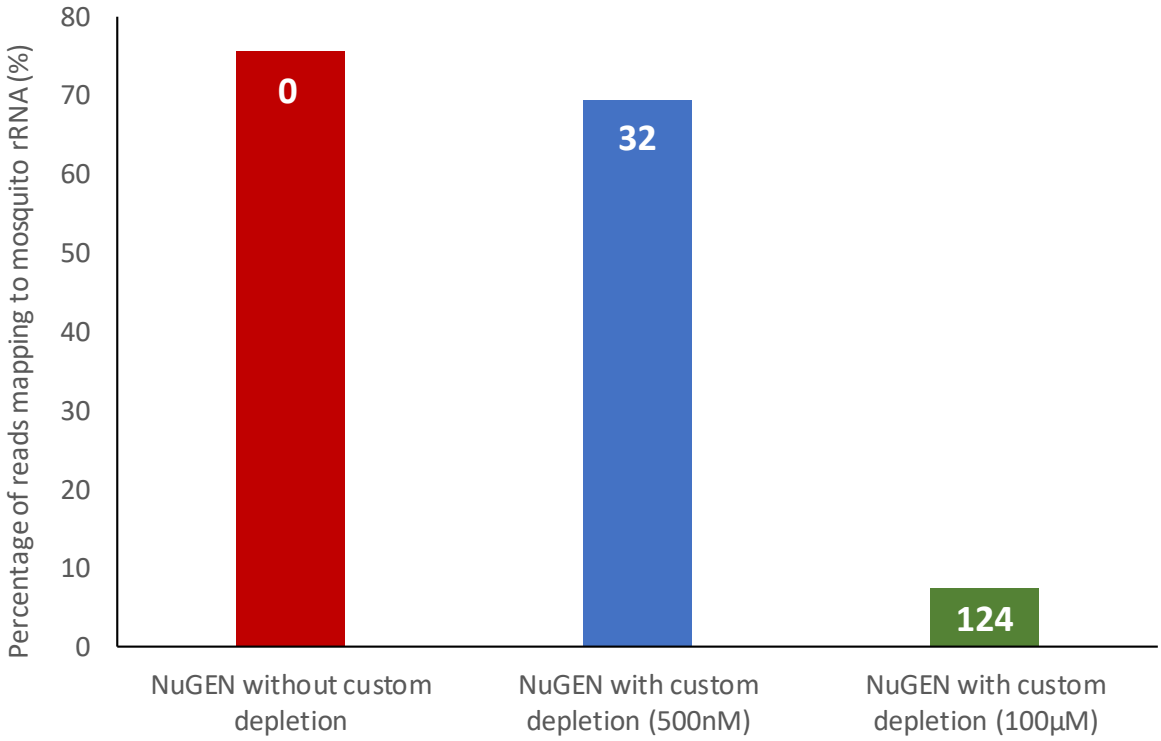


Fig. S2: Percentage of normalised reads mapping to mosquito ribosomal RNA (rRNA) from: **A)** A pool of 100 mosquitoes (99 *Culex australicus* and one *Aedes notoscriptus* infected with Ross River virus). NuGEN Ovation Universal RNA-Seq System library preparation was performed with combinations of undiluted (82.6 – 112 ng/µL) and diluted (30 ng/µL) input RNA, and customised rRNA depletion probe mixture added at the recommended 500 nM (blue) and 100 µM (red). The number of Ross River virus reads in each sample is displayed in white text. Reads for each sample were normalised to 600,000.



B) A pool of 1,000 mosquitoes (999 *Cx. australicus* and one *Ae. notoscriptus* infected with Ross River virus). NuGEN library preparation was performed without the customised rRNA depletion probe mixture (red), and with the customised rRNA depletion probe mixture added at the recommended concentration of 500 nM (blue) and at 100 µM (green). The number of Ross River virus reads in each sample is displayed in white text. Reads for each sample were normalised to 600,000.



C) Individual *Cx. quinquefasciatus*, *Cx. australicus*, and *Ae. notoscriptus* mosquitoes. Each sample was prepared with (red) and without (blue) customised rRNA depletion. The depletion was performed with customised probe mixture added at 100 µM. Reads for each sample were normalised to 800,000.

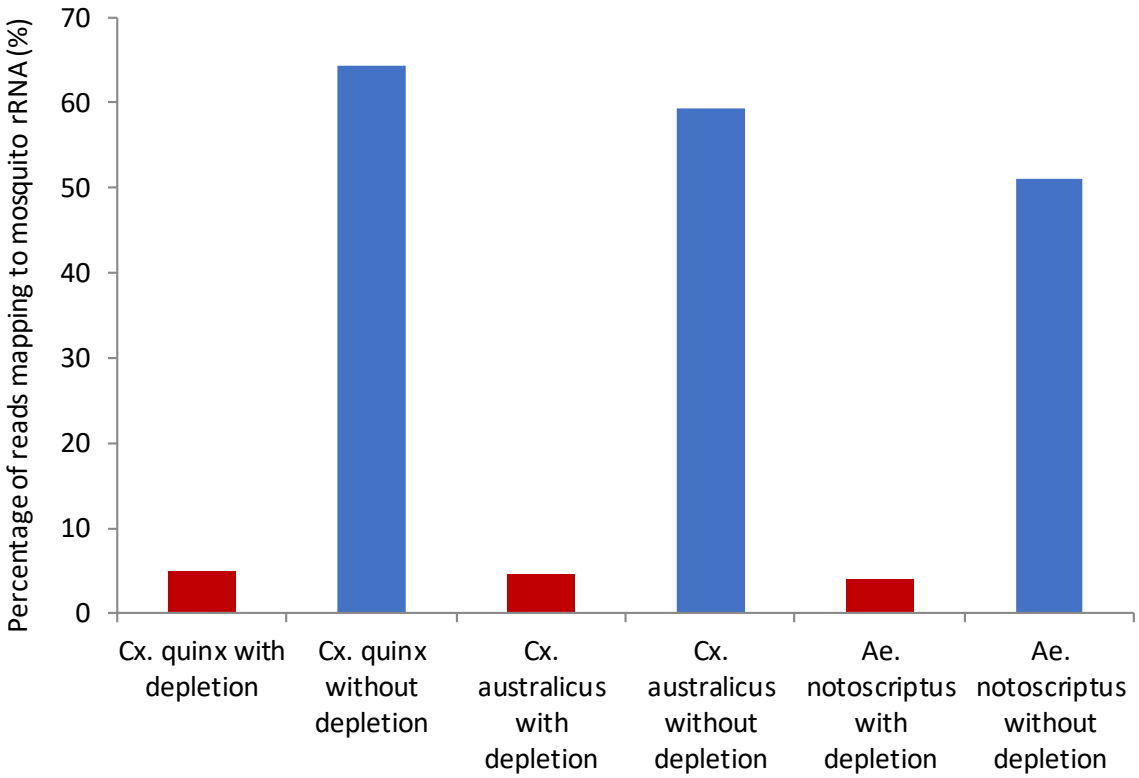
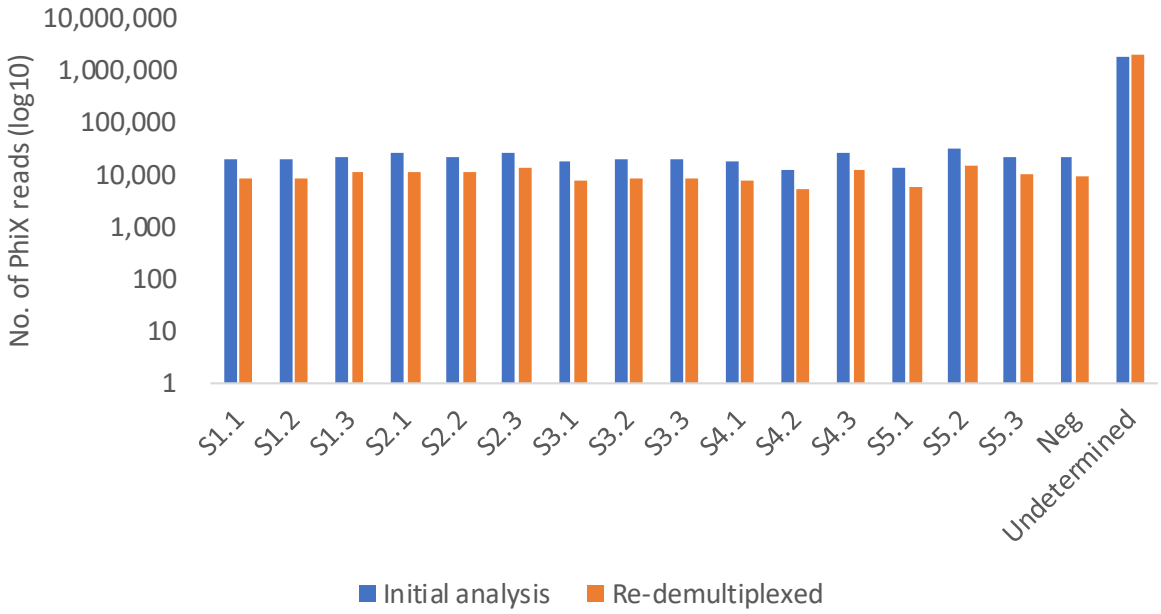


Fig. S3: Number of reads mapping to the PhiX genome from each set of FASTQ files output by the HiSeq sequencer. The initial read demultiplexing allowed one index mismatch, whereas the re-demultiplexed reads had no index mismatches.



CHAPTER 3

Enhanced arbovirus surveillance with high-throughput metatranscriptomic processing of field-collected mosquitoes

3.1 Chapter preface

The methods and positive detection criteria developed in Chapter 2 were evaluated using unsorted, bulk mosquito traps collected from different locations in Victoria, Australia as part of an existing arbovirus surveillance program in 2016-17. A curated database of Australian arboviruses was compiled and used to screen the metatranscriptomic trap data, with detections confirmed using established PCR methods. The detected arboviruses were analysed phylogenetically using whole genome or segment sequences assembled from the metatranscriptomic data. A curated database of cytochrome oxidase I (COI) sequences was also compiled to enable identification of mosquito and biting midge species from south-east Australia from the metatranscriptomic trap data. The metatranscriptomic arbovirus detection and mosquito identification was compared to corresponding surveillance data for the same traps produced by the arbovirus surveillance program in 2016-17. Further analysis of the metatranscriptomic data revealed a broad diversity of viruses endemic to mosquitoes, both known and previously undescribed. One of these novel viruses is further investigated in Chapter 4.

This chapter is in final stages of preparation, with intended journal submission to PLOS Neglected Tropical Diseases. The chapter is presented in a format compliant with the journal submission guidelines and includes supplementary information. Sequences generated in this chapter have been uploaded to GenBank and are awaiting approval. As such, the associated accession numbers have been kept blank and once assigned, will be added for journal submission.

3.2 Publication details

Title: Enhanced arbovirus surveillance with high-throughput metatranscriptomic processing of field-collected mosquitoes

Stage of publication: In preparation

Journal details: PLOS Neglected Tropical Diseases

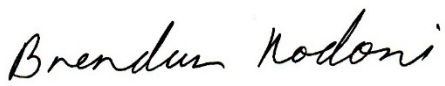
Authors: Jana Batovska, Peter T. Mee, Tim I. Sawbridge, Brendan C. Rodoni, and Stacey E. Lynch

3.3 Statement of joint authorship

All authors contributed to the conception and design of the experiment; JB performed the nucleic acid extractions, prepared the sequencing libraries, analysed the data, and wrote the manuscript; all authors contributed to the editing of the final manuscript and approved the version submitted for publication.

Statement from co-author confirming the contribution of the PhD candidate:

"As co-author of the manuscript 'Batovska, J., Mee, P.T., Sawbridge, T.I., Rodoni, B.C., and Lynch, S.E. (2020). Enhanced arbovirus surveillance with high-throughput metatranscriptomic processing of field-collected mosquitoes. PLOS Neglected Tropical Diseases, *in preparation*', I confirm that Jana Batovska has made the contributions listed above."



Professor Brendan Rodoni

22 June 2020

3.4 Manuscript

Enhanced arbovirus surveillance with high-throughput metatranscriptomic processing of field-collected mosquitoes

Batovska J[1,2], Mee PT[1], Sawbridge TI[1,2], Rodoni BC[1,2] and Lynch SE[1]

[1] Agriculture Victoria Research, AgriBio Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3083, Australia.

[2] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, 3086, Australia.

Keywords: Mosquitoes; Culicidae; arboviruses; surveillance; metatranscriptomics

Running title: Metatranscriptomics for arbovirus surveillance

Corresponding authors:

Jana Batovska

Email: jana.batovska@agriculture.vic.gov.au

Phone: +61 3 9032 7242

Stacey E. Lynch

Email: stacey.lynch@agriculture.vic.gov.au

Phone: +61 3 9032 7271

Abstract

Background

Surveillance programs are essential for the prevention and control of mosquito-borne arboviruses that cause serious diseases in humans and animals. Viral metatranscriptomic sequencing can enhance surveillance by enabling untargeted, high-throughput arbovirus detection. We used metatranscriptomic sequencing to screen field-collected mosquitoes for arboviruses to better understand how metatranscriptomics can be utilised in routine surveillance activities.

Methodology/Principal Findings

Following a significant flood event in 2016, more than 56,000 mosquitoes were collected over seven weeks from field traps set up in Victoria, Australia. The unsorted traps were split into samples of 1,000 mosquitoes or less and used for nucleic acid extraction. Libraries were prepared using a mosquito-specific ribosomal RNA depletion method and sequenced on the Illumina HiSeq. Screening of the metatranscriptomic data revealed the presence of five arboviruses relevant to public health: Ross River virus, Sindbis virus, Trubanaman virus, Umatilla virus, and Wongorr virus. A total of 33 arbovirus detections were made, 94% of which were confirmed using reverse transcription quantitative PCR (RT-qPCR). Analysis of cytochrome oxidase I (COI) sequences in the metatranscriptomic data also led to the detection of 12 mosquito and two biting midge species. Screening of mosquitoes from the same traps by an established public health arbovirus surveillance program corroborated the arbovirus and mosquito species detections made using metatranscriptomics. Whole genome or segment sequences assembled from the metatranscriptomic data were used to phylogenetically compare the detected arboviruses to previously sequenced strains. Further analysis of the viral sequences also provided the first look into the viral diversity of mosquitoes from south-east Australia, with the detection of 51 insect-specific viruses, both known and previously undescribed.

Conclusions/Significance

We have demonstrated the power of metatranscriptomics to enhance arbovirus surveillance by enabling untargeted arbovirus detection, providing genomic epidemiological data, and simultaneously identifying vector species from large, unsorted mosquito traps. The methods

and resources provided in this study will assist in the incorporation of metatranscriptomics into routine arbovirus surveillance activities.

Introduction

Arthropod-borne viruses (arboviruses) are distributed worldwide and in recent years have caused epidemics such as dengue, chikungunya, and Zika fever (Mayer et al., 2017). Dengue alone infects 390 million people a year, with a total economic cost of nearly US$40 billion (Selck et al., 2014). Almost 30% of emerging infectious diseases are arboviral, fuelled by increasing population growth, urbanisation, globalisation and international motility (Jones et al., 2008; Wilder-Smith et al., 2017). Arboviral infections can often be asymptomatic or present with non-specific symptoms, meaning that outbreaks can go undetected until containment is no longer feasible. For instance, clinical similarity to dengue and chikungunya viral infection enabled Zika virus (ZIKV) to circulate for over a year and half before the first detection in Brazil occurred in 2015, by which point it had already spread to over 40 countries (Grubaugh et al., 2019). Antibody cross-reactivity between flaviviruses and a lack of routine testing further hindered early detection (Priyamvada et al., 2016). Preparedness for these epidemics requires the ability to detect unexpected novel viral species and strains, and genomic information to reconstruct transmission dynamics and inform public health initiatives.

As the primary vector of arboviruses, mosquito populations are monitored by surveillance programs in order to detect and control arboviral activity. A common approach is to trap mosquitoes and test them for the presence of arboviruses using cell culture. This involves morphologically identifying the mosquitoes to species level and inoculating a subsample onto a range of suitable cell lines, which are then screened for arboviral presence using an appropriate immunological staining method or by observing for cytopathic effect (Knope et al., 2019; O'Brien et al., 2015). In recent years, molecular approaches such as reverse transcription PCR (RT-PCR) have been used for arbovirus detection, with a capacity to test pools containing thousands of mosquitoes (Ritchie et al., 2003). This is a significant upscale to cell culture, which loses sensitivity with pool sizes larger than 200 mosquitoes (Sutherland and Nasci, 2007), meaning that only small subsamples of trap catches are tested during flood seasons when thousands of mosquitoes are trapped each week (Knope et al., 2019). Due to the low arbovirus infection rates in mosquito populations, it is imperative to maximise sample sizes in order to increase detection probability (Gu and Novak, 2004). Although RT-PCR offers sensitivity and the ability to upscale surveillance, it requires *a priori* knowledge of the virus

sequence, which limits detection of divergent strains and restricts the discovery of unexpected novel viruses. Virus-specific PCRs also limit the number of targets and can decrease in sensitivity over time due to genomic drift in rapidly evolving viruses (Sozhamannan et al., 2015; Stellrecht, 2018).

Metatranscriptomics (total RNA sequencing) is an untargeted approach to virus detection that, unlike PCR, can generate whole genome sequences for all of the RNA viruses present in a sample. Focussing on RNA viruses is suited to arbovirus surveillance as all known arboviruses have an RNA genome, with the exception of African swine fever virus (ASFV), a double-stranded DNA virus that is transmitted by soft ticks (Gaudreault et al., 2020). The phylogenetic resolution offered by whole genome sequencing is particularly valuable in outbreak situations, where it can be used to reconstruct local virus transmission, elucidate the geographic origin of cases, track virus mutations, and identify highly transmissible strains (Pollett et al., 2020). For instance, genomics was used to uncover an unreported outbreak of ZIKV in Cuba and trace it to multiple introductions from other Caribbean islands, helping to direct vector control and further surveillance activities (Grubaugh et al., 2019). The untargeted nature of metatranscriptomics makes it ideal for arbovirus surveillance as it not only enables the detection of established viruses that can cause human disease, but also novel, unexpected viruses in mosquitoes, and other organisms of interest such as parasites (Ramos-Nino et al., 2020), bacteria, and fungi (Chandler et al., 2015). Furthermore, the mosquito species composition of the trap can be determined from the metatranscriptomic sequencing reads without manual sorting of the specimens (Belda et al., 2019), removing a major bottleneck in mosquito processing.

One of the challenges in implementing metatranscriptomics as a surveillance tool is the bioinformatics analysis involved in handling the sequencing data. A typical approach to metatranscriptomic virus detection involves comparing the sequencing reads or assembled contiguous sequences (contigs) to a reference database containing previously described virus sequences. Using a large, public database such as those offered by the National Center for Biotechnology Information (NCBI) provides comprehensive virus identification. However, these databases are littered with misannotated sequences, which can confound results (Steinegger and Salzberg, 2020). Alternatively, smaller, curated databases can be used to achieve more trustworthy results, but usually limit the scope of identification, leaving a

combined approach as the best but most time-consuming option (Chiu and Miller, 2019). In any circumstance, the method chosen needs to be tested and standardised in order to provide reliable and consistent results.

Decreases in sequencing costs have led to a rise in metatranscriptomic studies, with mosquito pools often sequenced to characterise viromes and discover new viruses (Pettersson et al., 2019; Sadeghi et al., 2018), investigate mosquito-specific virus ecology (Shi et al., 2017), and identify vector control candidates (Ramos-Nino et al., 2020; Zakrzewski et al., 2018). However, metatranscriptomics has yet to be applied to an established public health arbovirus surveillance program. As such, there is limited information on the sensitivity and specificity of metatranscriptomic arbovirus detection from mosquitoes and how this compares to established methods of arbovirus detection. A recent study introduced a positive detection criteria for metatranscriptomic detection of arboviruses from pooled mosquitoes (Batovska et al., 2019), but it is yet to be tested with traps containing diverse populations of mosquito and virus species.

The goal of this study is to assess the utility of metatranscriptomics in arbovirus surveillance. Using traps collected in 2016 following a significant flooding event in regional Victoria, Australia, we screened over 56,000 mosquitoes for arboviruses using metatranscriptomic sequencing. The results were confirmed using quantitative RT-PCR (RT-qPCR) and used to assess previously established positive detection criteria. The metatranscriptomic data was further utilised for mosquito and biting midge identification, and detection of insect-specific viruses, both known and previously undescribed. We also compared the metatranscriptomic results to existing surveillance program data for the same traps to examine the different outcomes of established methods and novel approaches to arbovirus surveillance.

Materials and methods

Mosquito collection and initial arbovirus screening

The mosquitoes used in this study were collected as part of the Department of Health and Human Services Victorian Arbovirus Disease Control Program (VADCP). Mosquito collection was performed on a weekly basis using carbon dioxide-baited encephalitis virus surveillance

(EVS) light traps (Rohe and Fall, 1979) overnight at three locations in Victoria, Australia (Figure 1). The three trapping locations were in the Rural City of Mildura (Mildura) (-34.249617, 142.218261), the Shire of Gannawarra (Gannawarra) (two traps spaced 2.2 km apart; -35.707128, 143.906764 and -35.720019, 143.925958), and the Wellington Shire Council (Wellington) (-38.206653, 147.396661). Traps were positioned in both bushland containing native animal hosts and near areas populated by humans in order to sample across enzootic and epizootic zones. The traps were collected over a seven week period during peak mosquito season from 7 November 2016 (Week 45) to 19 December 2016 (Week 51), following a significant flood event in the Murray-Darling Basin in September 2016 (Murray-Darling Basin Authority, 2016). Upon collection, the mosquitoes from the traps were immobilised at -20°C for 30 minutes, and then transported to the laboratory via chilled overnight delivery.

A subsample of mosquitoes was taken from each trap for arbovirus screening by the VADCP (Table 1). These mosquitoes were morphologically identified and screened for arboviruses using a cell culture-based system as previously described (Lynch et al., 2020). The remaining mosquitoes from each trap were stored at -20°C until used in this study for metatranscriptomic sequencing.

Sample preparation and nucleic acid extraction

Each mosquito trap was sorted into different sized subsamples. The first subsample for each trap consisted of 100 mosquitoes, which were counted and weighed. The remainder of each trap was then sorted into subsamples consisting of 1,000 mosquitoes, based on the weight of the initial 100 mosquitoes. This sample size was chosen based on previous findings indicating metatranscriptomic sequencing can detect an arbovirus from a single infected mosquito in a pool of 1,000 (Batovska et al., 2019). Any remaining mosquitoes were weighed and allocated as the last “remainder” subsample for each trap (estimated range: 78 to 906 mosquitoes). All mosquito subsamples were placed into 50 mL Falcon tubes and stored at -80°C until further use. A total of 86 subsamples from 21 traps were prepared.

To homogenise the mosquitoes for nucleic acid extraction, 3 mm glass beads (Sigma-Aldrich) and Buffer AVL (Qiagen) scaled according to mosquito number (Table S1) were added to each subsample. The mosquitoes were then mechanically homogenised for two 1 minute cycles at 1,200 rpm using a 2010 Geno/Grinder (SPEX SamplePrep). The homogenised mosquitoes

were centrifuged for 5 minutes at 15,344 x *g* and 140 µL of supernatant was removed from each subsample. Nucleic acid was extracted from the supernatant using the QIAamp Viral RNA Mini Kit (Qiagen) according to manufacturer's instruction, excluding the addition of carrier RNA. A double elution was performed using 2 x 40 µL of Buffer AVE and stored at -80°C until library preparation. An extraction negative control was included in each batch of extractions and consisted of 140 µL of Buffer AVL as input.

Metatranscriptomic sequencing

Libraries were prepared for metatranscriptomic sequencing using the NuGEN Ovation Universal RNA-Seq System with custom rRNA depletion. Briefly, 2 µL of undiluted RNA (68 – 200 ng) was DNase treated, converted into cDNA, and sheared into 200 – 400 bp fragments using a Covaris S220 Focused-ultrasonicator. A library negative control was included in each batch of library preparation and consisted of 2 µL of UltraPure water (Invitrogen) as input. Additionally, 2 µL of RNA from a pool of 1,000 mosquitoes containing a single Ross River virus (RRV) infected mosquito (Batovska et al., 2019) was used for a positive control library. After performing end repair, adaptor ligation and strand selection, the libraries were depleted of mosquito rRNA sequences using customised probes (Batovska et al., 2019). Libraries underwent 14 cycles of PCR amplification and were then purified using AMPure XP beads (Beckman Coulter). The libraries were quantified using a D1000 ScreenTape with the 2200 TapeStation (Agilent Technologies) and a dsDNA HS assay with the Qubit 3 Fluorometer (Thermo Fisher Scientific) and pooled into three samples of equimolar concentration. Each pooled sample was quantified using the TapeStation, diluted to 20 nM with UltraPure water, and treated with Free Adapter Blocking Reagent (Illumina) according to manufacturer's instructions in order to reduce index hopping (Illumina, 2017). The treated sample pools were quantified using the Qubit, diluted to 10 nM with 10 mM Tris buffer (pH 7.0; Invitrogen) and sequenced on a lane of the HiSeq 3000 (Illumina) with 2 x 150bp reads.

Read assembly and taxonomic classification

The reads were demultiplexed into subsamples and used for *de novo* assembly, performed using Trinity v2.4.0 (Grabherr et al., 2011) with read normalisation and trimming options selected. Assembled contigs over 500 bp (Belda et al., 2019) were taxonomically classified

using DIAMOND BLASTx v0.9.22.123 (Buchfink et al., 2015) with the NCBI non-redundant (nr) protein database (acquired 2nd September 2019) and an e-value threshold of 10^{-5}. Abundance was measured by mapping trimmed, interleaved reads back to the contigs using BWA-MEM v0.7.17-r1188 (Li, 2013) and obtaining read counts with SAMtools v1.9 idxstats (Li et al., 2009). Taxonomy, abundance and sample information was imported into RStudio v1.2.1335 (RStudio Team, 2015) for analysis and visualisation with phyloseq v1.28.0 (McMurdie and Holmes, 2013) and ggplot2 v3.2.1 (Wickham, 2009) packages. The abundance data was normalised to even sampling depth and taxa with a mean abundance lower than 10^{-5} were removed (Mariadassou et al., 2015). Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity was used to compare taxa in the trap subsamples, the positive control, and the extraction and library negative controls.

Metatranscriptomic mosquito and biting midge species identification

The mosquito (Diptera: Culicidae) and biting midge (Diptera: Ceratopogonidae) species in each trap were determined by using BLASTn v2.9.0+ (Camacho et al., 2009) to compare the subsample contigs to a custom database of cytochrome oxidase I (COI) barcode sequences. The database contained 138 COI sequences belonging to 29 mosquito species (Batovska et al., 2016) and 13 biting midge species (Dyce et al., 2007) found in Victoria, Australia. Biting midges were included in the database as they are often found in mosquito traps and can also transmit arboviruses (Tay et al., 2016). Members of the *Culex pipiens* mosquito species complex that cannot be differentiated by COI had their names conglomerated so that they would be counted as one (Batovska et al., 2016). The database is accessible via Figshare: https://doi.org/10.6084/m9.figshare.10246826.v3. The BLASTn was performed with an e-value threshold of 10^{-5} and the results were filtered for contigs with >300 bp alignment length and >95% identical match to COI sequences. Read counts for these contigs were acquired from the previously performed idxstats analysis, summed per species and per trap, and used to plot species abundance with ggplot2.

Targeted arbovirus screen

The metatranscriptomic data generated from the trap subsamples were screened for arboviruses of public health interest using a targeted custom database based on those listed

in Mackenzie et al. (1994) and Vasilakis et al. (2019). Whole genome sequences were used in the database if publicly available, and segmented genomes were merged so that each arbovirus was represented by a single sequence. The resulting database contains 74 arboviruses, representing nine viral families, and is available on Figshare: https://doi.org/10.6084/m9.figshare.12055830.v1. Screening was performed by mapping the subsample contigs and reads to the arbovirus database with BWA-MEM, and using BBMap pileup (Bushnell, 2017) to measure the average Fold Coverage by Reads (FCR), Percent Coverage by Contigs (%CC), and Percent Coverage by Reads (%CR). Each arbovirus coverage value in the subsamples was divided by any corresponding coverage in the negative controls, resulting in three coverage-based criteria: Fold Coverage by Reads ratio (FCR-r), Percent Coverage by Contigs ratio (%CC-r), and Percent Coverage by Reads ratio (%CR-r), with values ≥ 2 considered positive (Batovska et al., 2019). The performance of these criteria was compared by confirming every detection with RT-qPCR (as outlined below). Lastly, the SAMtools idxstats command was used to get read counts for the arboviruses in each trap based on the read alignments.

Confirmation of arbovirus detections using RT-qPCR

RT-qPCR was used to test all 86 subsamples for any arboviruses detected during the targeted arbovirus screen, with a total of five assays performed: Ross River virus (RRV), Sindbis virus (SINV), Trubanaman virus (TRUV), Umatilla virus (UMAV) and Wongorr virus (WGRV). The mosquito subsample RNA was diluted 1:5 with UltraPure water (Invitrogen) for use in the RT-qPCR assays. The RRV (Hall et al., 2011), SINV (Eshoo et al., 2007), TRUV (Lynch et al 2020) and UMAV (Cowled et al., 2009) primers were previously published, and the WGRV primers were designed in-house. Further details on the primers, as well as the PCR cycles and kits used, can be found in Table S2. For the probe-based RRV assay, subsamples with Ct <40 were considered positive; for the SYBR Green-based SINV, TRUV, UMAV and WGRV assays, subsample melt curves were also compared to negative and positive control melt curves to determine positivity.

Phylogenetic analysis of target arboviruses

In order to perform phylogenetic analysis, consensus sequences were derived from the read-based alignments that had >90% coverage of reference sequences from the targeted arbovirus screen. SAMtools mpileup and BCFtools consensus were used to generate the consensus sequences, and then MAFFT v7.429 (Katoh and Standley, 2013) was used to align them with other arbovirus sequences from GenBank. Full genome alignments were used for RRV and SINV (11,362 nt and 11,460 nt respectively), Segment M for TRUV (4,152 nt), and Segment 7 for UMAV (1,364 nt). Phylogenetic trees were created for each arbovirus alignment with PhyML v3.3 (Guindon et al., 2010) using maximum likelihood and a general time reversible (GTR) substitution model. Branch support was evaluated using 1,000 bootstrap replicates. The resulting trees were viewed and edited using Geneious v8.1.8 (Kearse et al., 2012). All arbovirus sequences generated for phylogenetic analysis have been uploaded to GenBank (acc. xxx-xxx).

Virome analysis

Contigs from the initial *de novo* assembly classified as viral by DIAMOND BLASTx were used to investigate the broader virome present in the trap subsamples. The contigs were compared to the NCBI nucleotide (nt) database (acquired 28th October 2019) using BLASTn with an e-value threshold of 10^{-5} and filtered to remove non-viral contigs. For the sake of brevity, the analysis was limited to RNA viruses by comparing the contigs to all viral RNA-dependent RNA polymerase (RdRp) protein sequences on RefSeq using DIAMOND BLASTx with an e-value threshold of 10^{-5}. Abundance estimates were determined by mapping reads back to the contigs and measuring read counts, as previously described, and summing the reads per trap. Viral contigs greater than 1,000 bp in length (Wille et al., 2019) had associated BLASTx and abundance information plotted in RStudio using ggplot.

Viral contig abundances were imported into phyloseq along with COI contig abundances (as previously identified via the COI BLASTn search). The read counts were normalised and filtered to remove taxa with a mean abundance below 10^{-5} (Mariadassou et al., 2015). The difference in viral and mosquito taxa amongst the subsamples was then visualised using an NMDS based on Bray-Curtis dissimilarity.

Results

Sample preparation and sequencing

Based on weight, a total of 62,218 mosquitoes were trapped in Gannawarra, Mildura and Wellington over a seven week period in 2016 (Weeks 45 – 51; Table 1; Figure 2). The VADCP subsampled 5,985 mosquitoes (9.6%) to use for cell culture-based arbovirus screening, with the remaining 56,233 mosquitoes (90.4%) used for metatranscriptomic sequencing. Mosquitoes were sorted into 2 – 15 subsamples per trap, resulting in a total of 86 mosquito subsamples for sequencing (Table 1). Additionally, there were three extraction negative controls, four library negative controls, and one positive control, resulting in a total of 94 samples for sequencing.

For the 86 mosquito subsamples, a mean of 10,575,201 paired reads were generated per subsample (range: 7,971,017 – 16,414,900). When the reads were mapped to taxonomically classified contigs (Figure 3), a mean of 78% (range: 17 – 99%) belonged to eukaryotes, the majority of which were arthropod species (mean: 70%; range: 12 – 98%). A substantial proportion of the reads (mean: 61%; range: 4 – 95%) were attributed to two ciliate species (*Oxytricha trifallax* and *Stylonychia lemnae*) and two nematode species (*Wuchereria bancrofti* and *Brugia timori*). However, further investigation indicated that these reads were derived from mosquito rRNA and so they were re-classified as arthropod for counting. Archaea and bacteria were represented by a mean of 1% of reads (range: 0.0 – 17%) and were not characterised as part of this study. The percentage of viral reads varied among the subsamples (1% – 82%), with certain traps having more viral reads than others. For instance, in Gannawarra the Week 45 and 46 traps had a mean of 55% viral reads, whereas the Week 47 – 51 traps had a mean of 8% viral reads.

The three extraction negative controls had a low number of sequencing reads compared to the subsamples, with a mean of 8,112 paired reads per sample (range: 5,249 – 10,139). Three out of the four library negative controls also had a low number of sequencing reads (mean: 112,252; range: 5,245 – 258,192), however there was one library negative control with 11,926,882 paired reads. Of the taxonomically classified reads for this library negative control,

80% were plant and 16% were bacterial, suggesting contamination. These plant and bacterial species were also present in some other samples from the same sequencing pool as the contaminated library negative control, but not in samples that were processed with it during library preparation, indicating the contamination occurred during sequencing. When comparing all taxa, both the extraction and library negative controls were distinct from the trap subsamples (Figure S1). The taxonomic composition of the positive control was similar to the subsamples.

Metatranscriptomic mosquito and biting midge species identification

The COI-based analysis of the metatranscriptomic data from all 86 subsamples identified 12 mosquito and two biting midge species in the 21 traps used in this study. The 12 mosquito species were detected over the seven-week period, with the two inland locations, Gannawarra and Mildura, sharing similar mosquito species in comparison to the coastal location, Wellington, which primarily had *Aedes camptorhynchus* (Figure 4A). The mosquito species composition changed over time, particularly in Gannawarra, where *Anopheles annulipes* and *Culex australicus/globocoxitus* populations were gradually replaced by *Culex annulirostris*. The two biting midge species were *Culicoides marksi* in Mildura, and *Culicoides multimaculatus* in Wellington, both detected at low abundances (<5% of mosquito and midge COI trap reads).

When only the 100-mosquito subsamples were used for COI-based metatranscriptomic analysis, abundance estimates for prevalent species were often comparable to those generated with all of the trap subsamples (Figure 4B). However, only 11 of the 14 species were present in the 100-mosquito subsamples, with low abundance (<5%) mosquito species *Aedes theobaldi* and *Tripteroides atripes*, and biting midge species *Culicoides marksi,* not detected. Analysis using only the 100-mosquito subsamples also resulted in the taxonomic dropout of other species, such as *Anopheles annulipes*, which was absent in seven of 19 traps positive for this species. In total, there were 26 taxonomic dropouts in 16 of the 21 traps when using only the 100-mosquito subsample data, compared to when all of the trap subsamples were used.

Assessment of positive detection criteria

Three criteria were assessed for arbovirus detection during the targeted screen of the metatranscriptomic data: Fold Coverage by Reads ratio (FCR-r), Percent Coverage by Contigs ratio (%CC-r), and Percent Coverage by Reads ratio (%CR-r), with values ≥ 2 considered positive (Batovska et al., 2019). Counting at trap level, a total of 9 detections were made using FCR-r, 15 using %CC-r, and 22 using %CR-r (Table S3A-C). No detections were made in the negative controls. All trap detections were confirmed using RT-qPCR, making %CR-r the most effective criterion to use for the targeted arbovirus screen of the metatranscriptomic data.

Arbovirus detection

Using the targeted database with the %CR-r criterion to screen the metatranscriptomic data resulted in the detection of five arboviruses: RRV, SINV, TRUV, UMAV and WGRV (Figure 5A). Out of the 86 subsamples, 25 were positive for one or more arbovirus, resulting in 33 detections. Counting at trap level, these represented 22 detections in 13 of the 21 traps. The majority of the trap detections came from Mildura (54.6%) and Gannawarra (31.8%), with only 13.6% from Wellington. RRV was detected in all three locations, SINV, TRUV and WGRV in Gannawarra and Mildura, and UMAV in Mildura. Based on read number per trap, the highest abundance arboviruses were SINV (mean 998.7 reads), TRUV (mean 663.8) and RRV (mean 514.5), while there were less reads attributed to UMAV (mean 137.7) and WGRV (mean 3). Out of the 33 detections, 22 (67%) were in a 1,000-mosquito subsample, nine (27%) in a remainder subsample (509 – 799 mosquitoes), and two (6%) in a 100-mosquito subsample, both of which were TRUV (Figure 5B).

Of the 33 metatranscriptomic subsample detections, 31 were confirmed using RT-qPCR (Table S4). The two unconfirmed detections were both SINV in Gannawarra from traps that had other subsamples positive for SINV via metatranscriptomics. The RT-qPCR testing revealed an additional 12 detections in the subsamples, or counting at trap level, an additional four detections: one RRV, two TRUV, and one WGRV (Table S4). Based on the RRV RT-qPCR results, lower Ct values corresponded to higher %CR (R^2 = 0.9, Figure S2). RRV-positive subsamples with a Ct <30 had %CR >98, whereas subsamples with a Ct >35 had %CR <10. A quarter (25.6%) of all the qPCR subsample detections had a Ct >35 (Table S4).

Comparison to routine surveillance program data

Despite representing a separate subsample of mosquitoes in each trap, the results of the initial screen performed by the VADCP in 2016 corresponded to the metatranscriptomic results (Figure 6). Relative abundance estimates based on morphological identification were similar to COI-based estimates for the most prevalent mosquito species (Figure 6A). However, there were differences in detection of lower abundance species, with morphology-based methods identifying 8 of the 14 species detected using metatranscriptomics. It should be noted that the VADCP does not identify or record biting midge species present in the traps, which accounts for two of the undetected species. Cell culture screening of the initial trap subsamples performed by the VADCP detected four out of the five arboviruses detected via metatranscriptomics (Figure 6B). UMAV was not detected, however, the VADCP orbivirus (UMAV and WGRV) screening was not as extensive as for other viruses of public health significance due to orbivirus isolates producing cytopathic effects in mosquito cells but not in mammalian cells (Lynch et al., 2020).

Phylogenetic analysis of arboviruses

Out of the 33 metatranscriptomic arbovirus detections made in the subsamples, only 10 provided the coverage required to generate a consensus sequence for use in phylogenetic analysis: three for RRV (whole genome, 11,362 nt); three for SINV (whole genome, 11,460 nt); three for TRUV (Segment M, 4,152 nt); and one for UMAV (Segment 7, 1,364 nt).

All three RRV detections grouped within the recently described Genotype 4 (G4) lineage (Figure 7A), to which all contemporary RRV strains belong (Michie et al., 2020). The Gannawarra and Mildura RRV detections were placed in the G4A sublineage, which contains mosquito-derived strains from Queensland (QLD) and Western Australia (WA), and human-derived strains from QLD. The Wellington RRV detection clustered with the smaller G4B sublineage, which consists of mosquito-derived strains from WA and human-derived strains from QLD. All three RRV detections shared >98% nucleotide identity with the G4 strains.

Phylogenetic analysis revealed that all three SINV detections belonged to the SINV-II genotype (Figure 7B), along with Australian strain 18953, which was isolated from *Culex annulirostris* mosquitoes in 1975, and Chinese strain YN_222, which was isolated from a midge in 2013

(Pickering et al., 2019). SINV from both Gannawarra and Mildura shared 96% nucleotide identity with Australian strain 18953, and 90% with Chinese strain YN-222. The other Australian SINV strain, SW6562, which was isolated in 1984 and belongs to the SINV-VI or the south-west Australia genotype (Saleh et al., 2003), shared 72% nucleotide identity with SINV from the Gannawarra and Mildura traps. Interestingly, SINV from one of the Gannawarra subsamples shared more nucleotide identity with SINV from the Mildura subsample (99.51%) than with SINV from the other Gannawarra subsample (99.16%), despite belonging to the same trap.

When compared with Segment M sequences from orthobunyaviruses in the Mapputta group, the three TRUV detections clustered with other TRUV strains (Figure 7C), forming a clade with the type strain MRM3630, isolated in QLD in 1965, strain SW27572, isolated in WA in 1993, and strain Murrumbidgee 934 (also known as Murrumbidgee virus), isolated in New South Wales (NSW) in 1997 (Briese et al., 2016; Coffey et al., 2014; Shchetinin et al., 2015). These three TRUV strains were all isolated from *Anopheles annulipes* mosquitoes. The TRUV detections from Gannawarra in Week 46 and Mildura in Week 51 were similar (99.47%) and shared >98% nucleotide identity with strains SW27572 and Murrumbidgee 934. The TRUV from Mildura in Week 47 was 5% different to the other two detections, sharing most nucleotide identity with strain MRM3630 (97.3%).

Of the three UMAV detections from Mildura, only the one from Week 50 had enough coverage of a genome segment to allow phylogenetic analysis (Figure 7D). Comparison of Segment 7 showed the Mildura UMAV detection was most similar to UMAV M4941_15 (94.19% nucleotide identity), which was isolated from *Culex quinquefasciatus* mosquitoes in 2015, also in Victoria (Batovska et al., 2019). The two Victorian strains grouped with the Japanese Koyama Hill virus (KHV) (Ejiri et al., 2014), forming a separate clade to the two American UMAV strains.

Virome ecology

In addition to the five arboviruses detected using the targeted database, contigs matching 51 other viruses were assembled from the trap subsample reads, ranging in size from 1 – 20 kbp (Figure 8). Of the 51 viruses, 32 are from an existing viral group, with a total of 16 viral families or orders represented, whereas the other 19 viruses are currently unclassified. Some of the

viruses are closely related (Figure S3) and likely belong to the same viral species. All of the viruses are insect-specific or have no known vertebrate host, except for Hypsignathus monstrosus dicistrovirus (HMDV), which was sequenced from fruit bats but originated in arthropods (Bennett et al., 2019), and Fisavirus 1 (FSV1), which was sequenced from the intestinal content of freshwater carp and is also of arthropod origin (Reuter et al., 2015). Furthermore, the contigs matching HMDV and FSV1 had a mean 50% and 38% RdRp protein similarity to the reference sequences, respectively, which is indicative of new species (Wille et al., 2019). Almost half of the viruses (25 out of 51) shared <90% RdRp protein similarity with the matching contigs and likely represent novel viral taxa. The longest contig for each virus has been uploaded to GenBank (acc. xxx-xxx).

Based on read number per positive trap, the most abundant viruses were those matching Hubei arthropod virus 1 (mean 2,775,315.8 reads), Ngewotan virus (1,063,130.1), Culex-associated Tombus-like virus (419,002.7), and Yongsan picorna-like virus 2 (313,737.5). Certain viruses fluctuated in abundance over time, such as Ngewotan virus (Colmant et al., 2020), which in Gannawarra went from a mean 2,567,730 reads in Weeks 45 to 48, to a mean 34,499 reads in Weeks 49 to 51. This pattern of abundance for Ngewotan virus was repeated in Mildura but not in Wellington. Other viruses also had location-specific patterns of abundance, with 15 of the 51 viruses detected in only one of the three locations (Figure 8).

Differences in virus abundance and geography were influenced by the mosquito species present in the traps. For instance, the pattern of abundance for Ngewotan virus (Figure 8) resembles that of *Culex australicus/globocoxitus* (Figure 4A). Additionally, the presence of viruses such as Aedes camptorhynchus reo-like virus in only Wellington can be related to the abundance of *Aedes camptorhynchus* mosquitoes in traps from that location. When the virus and mosquito species were compared in the three locations, the species found in inland Gannawarra and Mildura were similar compared to the species found in coastal Wellington (Figure 9). Furthermore, the virus and mosquito species found in Gannawarra and Mildura varied over time, however in Wellington there was minimal variation in the species detected over the seven week trapping period. For each trap, the species composition was largely similar amongst the subsamples, with outliers often a result of mosquito number variations (Figure S4).

Discussion

We demonstrate the utility of metatranscriptomics as a high-throughput arbovirus surveillance tool by screening over 56,000 mosquitoes from 21 traps and detecting five arboviruses of public health interest. Additionally, the metatranscriptomic data was used to determine the species composition of the traps and to survey the broader viral diversity, highlighting the versatility of the data.

The metatranscriptomic COI analysis resulted in the detection of 12 mosquito species and two biting midge species in the traps (Figure 4A). Whilst COI could not differentiate certain species (i.e. members of the *Culex pipiens* complex), it detected others that are difficult to identify morphologically, such as *Culex palpalis*, which is almost indistinguishable from the closely related *Culex annulirostris* (Jansen et al., 2013). Furthermore, by detecting biting midge species such as *Culicoides marksi*, a vector of animal arboviruses (Tay et al., 2016), metatranscriptomics further extends the utility of surveillance to veterinary health. Whereas morphological identification of a separate insect family requires extensive taxonomical expertise, metatranscriptomic identification only requires the addition of reference sequences to the database used during analysis, thereby enabling the detection of any species with a distinct, curated barcode sequence. Continued efforts to grow comprehensive barcode databases based on accurately identified specimens for species relevant to surveillance are essential in broadening the capacity of metatranscriptomic species identification (Ekrem et al., 2007) (Table 2).

Targeted screening of the metatranscriptomic data resulted in the detection of five arboviruses relevant to public health: RRV, which causes notifiable disease in humans and animals (Harley et al., 2001); SINV and TRUV, which have been linked to human and animal arboviral infection (Boughton et al., 1990; Doherty et al., 1970; Saleh et al., 2003); and UMAV and WGRV, which are both serologically linked to infections in animals (Belaganahalli et al., 2011; Doherty et al., 1973). The coverage provided by metatranscriptomic sequencing enabled phylogenetic analysis using long stretches of the viral genomes (range: 1,364 – 11,460 nt), offering valuable insights on genotypic diversity, viral lineages, and geographical differences. For instance, whole genome phylogenies revealed the presence of two

geographically separated RRV sublineages in Victoria (Figure 7A), with the inland Gannawarra and Mildura RRV detections placed within the G4A sublineage, and the coastal Wellington RRV detection placed within the G4B sublineage (Michie et al., 2020). The distinct lineages warrant monitoring as ongoing evolution may lead to changes in RRV fitness and virulence (Coffey et al., 2013). As for SINV, it is unclear why the detections in two subsamples from the same Gannawarra trap were less similar than the SINV detected in Mildura (Figure 7B), reflecting a need for more extensive sampling to better understand the genetic diversity of SINV in Victoria. However, this demonstrates how splitting traps into subsamples for screening helps to uncover the viral diversity present within trap mosquito populations. We have made available all of the arbovirus sequences used for phylogenetic analysis, which will strengthen future analyses.

More publicly available reference material could have improved the analysis of the detected orbiviruses (UMAV and WGRV). Phylogenetic analysis of the NS2 protein (Segment 7) from UMAV showed that it was most similar to another UMAV sequence from Victoria (Batovska et al., 2019) and KHV from Japan (Ejiri et al., 2014), together forming a clade separate to two UMAV strains from the USA. Though useful to know these relationships, it is a limited representation of the *Umatilla virus* species, which also contains Stretch Lagoon orbivirus (SLOV), Minnal virus (MINV), Netivot virus (NETV), and Llano Seco virus (LLSV) (Belaganahalli et al., 2011). These viruses do not have available reference sequences, apart from SLOV, which has sequences for all segments except for 3, 7, and 10 (Cowled et al., 2009; Tangudu et al., 2019). The issue of limited reference material was particularly evident with WGRV, which only had 1,368 bp of genomic sequence available for inclusion in the Australian arbovirus reference database used to screen the trap subsamples (https://doi.org/10.6084/m9.figshare.12055830.v1) and insufficient coverage (<30%) to perform phylogenetic analysis. In an attempt to acquire more of the WGRV genome, the assembled contigs were screened for sequences similar to those publicly available for WGRV but none were found (data not shown). The search was confounded by the segmented nature of the WGRV genome and the presence of other orbiviruses in the same traps (Figures 5A and 9). Evidently the utility of metatranscriptomic surveillance is dependent on the availability of reference genomic material, stressing the importance of sequencing archival, curated arbovirus collections (Pyke and Warrilow, 2016) (Table 2).

The majority of metatranscriptomic arbovirus detections were confirmed by RT-qPCR, with only two subsample detections missed by the SINV assay (Table S4). The four metatranscriptomic SINV subsample detections that were confirmed by RT-qPCR all had high Ct values (>38), even though three of them had high %CR values (>98) via metatranscriptomic sequencing, which typically corresponds to low Ct values (Figure S2). It is likely primer inefficiency is responsible for the lower sensitivity and high Ct values of the SINV PCR assay. The SINV primers used were designed to detect a broad range of alphavirus species and only the SINV-I genotype was included in the primer design (Eshoo et al., 2007), possibly missing important differences in the primer region present in the SINV-II genotype that was detected in the trap subsamples. On inspection, there were four and six mismatches within the 24 and 28 base SINV primers used, respectively, when compared to the SINV genomes assembled from the subsample metatranscriptomic data (data not shown). Primers that are designed to amplify a broad range of targets are known to have lower detection sensitivity (Lemmon and Gardner, 2008), especially when dealing with complex sample types such as bulk homogenised mosquitoes, which can contain PCR inhibitors (Schrader et al., 2012). The untargeted approach of metatranscriptomics means differences in arbovirus genotypes do not affect detection, providing enhanced surveillance capabilities.

Apart from SINV detection, screening with RT-qPCR offered highly sensitive results, with an additional four arbovirus detections made at a trap level: 1 RRV, 2 TRUV and 1 WGRV. The RRV detection had a Ct value of 39.82, indicating RRV was present at a very low concentration and at the limit of detection (Hall et al., 2011). These results are consistent with other studies that have shown RT-qPCR is more sensitive than metatranscriptomics in detecting specific viruses from complex sample types (Bibby and Peccia, 2013; Fernandez-Cassi et al., 2017, 2018). However, it is misleading to compare the two technologies on only this measure, considering that metatranscriptomics also detects an abundance of other viruses, and also the host species, all within the single sequencing reaction. As for TRUV, one of the qPCR detections was also detected via metatranscriptomics but had a %CR of 1.52, which fell below the positive detection threshold of ≥2. This was true of two other subsample detections (Table S4) and may be reason to further optimise the positive detection threshold of %CR for metatranscriptomic detection of viruses. Finally, the missed WGRV detection is likely affected by the lack of whole genome reference sequences, as discussed previously. The use of short

sequences for screening lowers the chance of detection, particularly if the virus is of low abundance, highly divergent, or has a segmented genome (Schlaberg et al., 2017). The utility of genomic surveillance will increase over time as more whole genome sequences of local arboviruses are made available (Table 2).

The metatranscriptomic results were supported by the initial screening performed by the VADCP during the 2016-17 surveillance season, which recorded similar mosquito species and arbovirus detections, despite using a separate subsample of mosquitoes in each trap (Figure 6). The use of labour-intensive techniques, such as morphological mosquito identification and cell-culture based arbovirus detection, meant that during the surveillance season the VADCP screened just under 6,000 mosquitoes (9.6%) from the 21 traps used in this study, which contained over 62,000 mosquitoes in total. The high-throughput capabilities of metatranscriptomics allowed screening of the other 90.4% of mosquitoes, amounting to over 56,000 mosquitoes in total, making this the largest mosquito metatranscriptomic study to date. In addition to mosquito species identification and arbovirus detection, metatranscriptomics was able to provide additional surveillance information, including biting midge identification, genomic information for arboviruses, and a profile of the trap virome, all within a single reaction. The metatranscriptomic laboratory protocol required one technician and took approximately seven days: three days for sample preparation and four days for sequencing using the Illumina HiSeq. The processing time could be considerably reduced by automating parts of the sample preparation and utilising a different sequencer, such as the Illumina NovaSeq, which halves the sequencing time to two days (Piper et al., 2019). The cost to metatranscriptomically screen each subsample containing up to 1,000 mosquitoes was approximately AUD$230, not including labour. Switching to in-house rRNA depletion and the NovaSeq would reduce this to approximately AUD$110 (Table 2). As high-throughput sequencing continues to decrease in price and turnaround time, metatranscriptomics will progressively become the most cost-effective option for arbovirus surveillance.

The subsampling employed by the VADCP in response to surges in mosquito numbers during the 2016-17 surveillance season (Table 1) assumes that the mosquitoes and arboviruses present in one subsample will be indicative of the whole trap. To investigate if this was true using metatranscriptomic sequencing, an additional analysis using only 100-mosquito

588 subsamples from each trap was performed. While the COI-based species composition was
589 predominantly the same as when all trap subsamples were used (Figure 4B), the arbovirus
590 detections were greatly reduced, with only two metatranscriptomic arbovirus detections
591 made using the 100-mosquito subsamples, compared to 33 detections made using all of the
592 trap subsamples (Figure 5B). These results are reflective of the low arbovirus infection rates
593 in mosquito populations and supports sequencing all trapped mosquitoes to maximise the
594 probability of detection (Gu and Novak, 2004) (Table 2). Future research could investigate the
595 effect of homogenising the whole trap and subsampling this for surveillance.

596 Metatranscriptomic arbovirus detection was determined by the percent coverage of the
597 arbovirus genome by reads ratio (%CR-r), which was shown to be the most effective criterion
598 for detection from the field traps (Table S3). This contrasts previous research that used the
599 percent genome coverage by contigs ratio (%CC-r) and average fold genome coverage by
600 reads ratio (FCR-r) as criteria for positive detection of arboviruses, based on pooled mosquito
601 samples spiked with known concentrations of arboviruses (Batovska et al., 2019). The
602 differences in criteria performance may partly be explained by the low abundance of
603 arboviruses in some of the field traps (Figure 4A), compared to the high titres used in the
604 spiking study. Differences in abundance may have also been due to the degradation of viral
605 RNA in the field traps, compared to the viral spikes grown via cell culture (Kramer et al., 2001;
606 Mavale et al., 2012). Low arbovirus abundance hindered contig assembly and prevented
607 depth of coverage, thereby lowering the sensitivity of %CC-r and FCR-r, respectively,
608 rendering them less useful when applied to field traps (Table S3). Unlike the spiking study, the
609 low arbovirus abundance also meant there were no arbovirus reads in the negative controls
610 as a result of index cross-talk (Larsson et al., 2018), thereby negating the need for
611 normalisation of arbovirus coverage metrics and changing the efficacy of the criteria. As such,
612 the higher sensitivity offered by the %CR-r criterion makes it the most suitable for arbovirus
613 detection from field traps. Further testing of the criterion using statistically robust methods
614 is encouraged to validate its usage in routine surveillance activities (Schlaberg et al., 2017)
615 (Table 2).

616 Further analysis of the metatranscriptomic trap data revealed a broad viral diversity, with
617 assembled contigs matching 51 viruses specific to or originating from arthropods (Figure 8).
618 This is the first time the viral diversity of mosquitoes from south-east Australia has been

explored using a metatranscriptomic approach. Based on amino acid similarity (range: 31.3–100%), some of the detected viruses are novel and warrant further investigation. Insect-specific viruses can assist in understanding virus evolution (Li et al., 2015), be applied to vaccine production and diagnostics (Hobson-Peters et al., 2019), and have potential as biocontrol agents (Öhlund et al., 2019). Future efforts in these areas will be supported by the representative viral sequences provided by this study (acc xxx-xxx). These sequences can also assist routine surveillance activities by enabling identification of sequences belonging to the endemic mosquito virome and reducing viral dark marker, which improves the efficiency of detecting unexpected or emerging viruses (Batson et al., 2020).

While the exploration of broader viral diversity in the field traps enabled the detection of both known and novel viruses, it was also a complex and time-consuming process that would challenge routine surveillance activities. Unlike the targeted arbovirus screening of the metatranscriptomic data, there is no established criteria for the positive detection of previously undescribed viruses. While there are standards for reporting metatranscriptomic virus genomes (Roux et al., 2019), it is difficult to classify divergent viruses that often only have partially assembled genomes, let alone establish robust detection criteria. Determining whether a contig sequence is divergent enough from known viruses to constitute a new taxon is dependent on guidelines that vary between different viral groups (Simmonds, 2015). This is further complicated by a lack of formal taxonomic classification for many of the viruses derived from metatranscriptomic sequencing. Often these unclassified viruses have been sequenced from invertebrate samples and are the closest match to viral sequences generated from mosquito samples (Figure 8). Unclassified viruses hinder the efficiency of metatranscriptomic arbovirus surveillance because they require further investigation to determine if their detection is of significance to public health, which typically involves lengthy phylogenetic analyses (Dacheux et al., 2014; Harvey et al., 2019). Metatranscriptomic arbovirus surveillance would benefit from the development of an analysis tool that would automate the process of determining the public health risk associated with novel or unclassified viral sequences detected in field traps (Table 2).

By processing unsorted, bulk mosquito traps for untargeted, high-throughput arbovirus detection and vector species identification, we have demonstrated metatranscriptomics as a high-value resource for arbovirus surveillance programs. The methods and resources

650 presented here, including the curated reference sequence databases and refined positive
651 detection criteria, can help facilitate the incorporation of metatranscriptomics into routine
652 surveillance activities. Future efforts should focus on standardising operating procedures,
653 further refining limits of detection to diagnostic standards, optimising the protocol to lower
654 assay cost, developing user-friendly data analysis software, and expanding reference
655 sequence databases (Table 2). The implementation of mosquito-based metatranscriptomic
656 arbovirus detection will herald a new era of genomic surveillance that strengthens our ability
657 to detect, track, and contain arboviral outbreaks and improve public health.

658

659 **Data availability**

660 The unprocessed Illumina HiSeq FASTQ read files for all 86 trap subsamples have been
661 deposited into the NCBI SRA database (BioProject ID PRJNA642916). The custom nucleotide
662 database used for the COI-based identification of mosquito and biting midge species is
663 available on Figshare (https://doi.org/10.6084/m9.figshare.10246826.v3), as is the Australian
664 arbovirus database (https://doi.org/10.6084/m9.figshare.12055830.v1). The RRV, SINV,
665 TRUV and UMAV sequences used for phylogenetic analysis are available on GenBank (acc xxx-
666 xxx), as are the longest contig sequences assembled for each virus detected as part of the
667 broader virome analysis (acc xxx-xxx).

668

669 **Author contributions**

670 JB: Conceptualisation, data curation, formal analysis, investigation, methodology, project
671 administration, resources, validation, visualisation, writing – original draft.

672 PTM: Methodology, resources, supervision, writing – review & editing.

673 TIS: Methodology, supervision, writing – review & editing.

674 BCR: Methodology, supervision, writing – review & editing.

675 SEL: Conceptualisation, funding acquisition, methodology, project administration, resources,
676 supervision, writing – review & editing.

Acknowledgements

The authors would like to thank Barry Curtain, Dale Hutchinson, Graeme Murray, Joel Bray, Russell Barnes and Shane Leerson for performing mosquito trapping associated with this study; Karen Brown for assisting with the counting, weighing and sorting of the trap subsamples, and for morphologically identifying the mosquitoes screened as part of the VADCP; Shani Wong and Jodie Crowder for performing the viral screening as part of the VADCP; and Jonathan Darbro and Elise Kho for providing the RRV-infected mosquito used in the positive control sample.

The VADCP is funded by the Department of Health and Human Services, this study was funded by the Agriculture Victoria Research Innovation Fund through the Department of Jobs, Precincts and Regions, and J.B. is supported by an Australian Government Research Training Program Scholarship and an Agriculture Victoria Research Top-Up Scholarship.

References

Batovska, J., Blacket, M.J., Brown, K., and Lynch, S.E. (2016). Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. Ecology and Evolution *6*, 3001–3011.

Batovska, J., Mee, P.T., Lynch, S.E., Sawbridge, T.I., and Rodoni, B.C. (2019). Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool. Scientific Reports *9*, 1–13.

Batson, J., Dudas, G., Haas-Stapleton, E., Kistler, A.L., Li, L.M., Logan, P., Ratnasiri, K., and Retallack, H. (2020). Single mosquito metatranscriptomics recovers mosquito species, blood meal sources, and microbial cargo, including viral dark matter. BioRxiv 2020.02.10.942854.

Belaganahalli, M.N., Maan, S., Maan, N.S., Tesh, R., Attoui, H., and Mertens, P.P.C. (2011). Umatilla virus genome sequencing and phylogenetic analysis: identification of Stretch Lagoon orbivirus as a new member of the *Umatilla virus* species. PLOS ONE *6*, e23605.

Belda, E., Nanfack-Minkeu, F., Eiglmeier, K., Carissimo, G., Holm, I., Diallo, M., Diallo, D., Vantaux, A., Kim, S., Sharakhov, I.V., et al. (2019). De novo profiling of RNA viruses in *Anopheles* malaria vector mosquitoes from forest ecological zones in Senegal and Cambodia. BMC Genomics *20*, 664.

Bennett, A.J., Bushmaker, T., Cameron, K., Ondzie, A., Niama, F.R., Parra, H.-J., Mombouli, J.-V., Olson, S.H., Munster, V.J., and Goldberg, T.L. (2019). Diverse RNA viruses of arthropod

origin in the blood of fruit bats suggest a link between bat and arthropod viromes. Virology *528*, 64–72.

Bibby, K., and Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environmental Science & Technology *47*, 1945–1951.

Boughton, C.R., Hawkes, R.A., and Naim, H.M. (1990). Arbovirus infection in humans in NSW: seroprevalence and pathogenicity of certain Australian bunyaviruses. Australian and New Zealand Journal of Medicine *20*, 51–55.

Briese, T., Williams, D.T., Kapoor, V., Diviney, S.M., Certoma, A., Wang, J., Johansen, C.A., Chowdhary, R., Mackenzie, J.S., and Lipkin, W.I. (2016). Analysis of arbovirus isolates from Australia identifies novel bunyaviruses including a Mapputta group virus from Western Australia that links Gan Gan and Maprik viruses. PLOS ONE *11*, e0164868.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods *12*, 59–60.

Bushnell, B. (2017). BBMap short read aligner, and other bioinformatic tools. (https://sourceforge.net/projects/bbmap/).

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.

Chandler, J.A., Liu, R.M., and Bennett, S.N. (2015). RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. Frontiers in Microbiology *6*, 185.

Chiu, C.Y., and Miller, S.A. (2019). Clinical metagenomics. Nature Reviews Genetics *20*, 341–355.

Coffey, L.L., Forrester, N., Tsetsarkin, K., Vasilakis, N., and Weaver, S.C. (2013). Factors shaping the adaptive landscape for arboviruses: implications for the emergence of disease. Future Microbiology *8*, 155–176.

Coffey, L.L., Page, B.L., Greninger, A.L., Herring, B.L., Russell, R.C., Doggett, S.L., Haniotis, J., Wang, C., Deng, X., and Delwart, E.L. (2014). Enhanced arbovirus surveillance with deep sequencing: identification of novel rhabdoviruses and bunyaviruses in Australian mosquitoes. Virology *448*.

Colmant, A.M.G., O'Brien, C.A., Newton, N.D., Watterson, D., Hardy, J., Coulibaly, F., Bielefeldt-Ohmann, H., Warrilow, D., Huang, B., Paramitha, D., et al. (2020). Novel monoclonal antibodies against Australian strains of negeviruses and insights into virus structure, replication and host-restriction. The Journal of General Virology *101*, 440–452.

Cowled, C., Palacios, G., Melville, L., Weir, R., Walsh, S., Davis, S., Gubala, A., Lipkin, W.I., Briese, T., and Boyle, D. (2009). Genetic and epidemiological characterization of Stretch Lagoon orbivirus, a novel orbivirus isolated from *Culex* and *Aedes* mosquitoes in northern Australia. The Journal of General Virology *90*, 1433–1439.

746 Dacheux, L., Cervantes-Gonzalez, M., Guigon, G., Thiberge, J.-M., Vandenbogaert, M.,
747 Maufrais, C., Caro, V., and Bourhy, H. (2014). A preliminary study of viral metagenomics of
748 French bat species in contact with humans: identification of new mammalian viruses. PLOS
749 ONE *9*, e87194.

750 Doherty, R.L., Whitehead, R.H., Wetters, E.J., Gorman, B.M., and Carley, J.G. (1970). A survey
751 of antibody to 10 arboviruses (Koongol group, Mapputta group and ungrouped) isolated in
752 Queensland. Transactions of the Royal Society of Tropical Medicine and Hygiene *64*, 748–753.

753 Doherty, R.L., Carley, J.G., Standfast, H.A., Dyce, A.L., Kay, B.H., and Snowdon, W.A. (1973).
754 Isolation of arboviruses from mosquitoes, biting midges, sandflies and vertebrates collected
755 in Queensland, 1969 and 1970. Transactions of the Royal Society of Tropical Medicine and
756 Hygiene *67*, 536–543.

757 Dyce, A.L., Bellis, G.A., and Muller, M.J. (2007). Pictorial Atlas of Australasian *Culicoides* Wings
758 (Diptera: Ceratopogonidae) (Australian Biological Resources Study).

759 Ejiri, H., Kuwata, R., Tsuda, Y., Sasaki, T., Kobayashi, M., Sato, Y., Sawabe, K., and Isawa, H.
760 (2014). First isolation and characterization of a mosquito-borne orbivirus belonging to the
761 species *Umatilla virus* in East Asia. Archives of Virology *159*, 2675–2685.

762 Ekrem, T., Willassen, E., and Stur, E. (2007). A comprehensive DNA sequence library is
763 essential for identification with DNA barcodes. Molecular Phylogenetics and Evolution *43*,
764 530–542.

765 Eshoo, M.W., Whitehouse, C.A., Zoll, S.T., Massire, C., Pennella, T.-T.D., Blyn, L.B., Sampath,
766 R., Hall, T.A., Ecker, J.A., Desai, A., et al. (2007). Direct broad-range detection of alphaviruses
767 in mosquito extracts. Virology *368*, 286–295.

768 Fernandez-Cassi, X., Timoneda, N., Gonzales-Gustavson, E., Abril, J.F., Bofill-Mas, S., and
769 Girones, R. (2017). A metagenomic assessment of viral contamination on fresh parsley plants
770 irrigated with fecally tainted river water. International Journal of Food Microbiology *257*, 80–
771 90.

772 Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J.,
773 Figuerola, N., Bofill-Mas, S., Abril, J.F., and Girones, R. (2018). Metagenomics for the study of
774 viruses in urban sewage as a tool for public health surveillance. The Science of the Total
775 Environment *618*, 870–880.

776 Gaudreault, N.N., Madden, D.W., Wilson, W.C., Trujillo, J.D., and Richt, J.A. (2020). African
777 swine fever virus: an emerging DNA arbovirus. Frontiers in Veterinary Science *7*.

778 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan,
779 L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-
780 Seq data without a reference genome. Nature Biotechnology *29*, 644–652.

781 Grubaugh, N.D., Saraf, S., Gangavarapu, K., Watts, A., Tan, A.L., Oidtman, R.J., Ladner, J.T.,
782 Oliveira, G., Matteson, N.L., Kraemer, M.U.G., et al. (2019). Travel surveillance and genomics
783 uncover a hidden Zika outbreak during the waning epidemic. Cell *178*, 1057-1071.e11.

784 Gu, W., and Novak, R.J. (2004). Short report: detection probability of arbovirus infection in
785 mosquito populations. The American Journal of Tropical Medicine and Hygiene *71*, 636–638.

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

789 Hall, R., Prow, N., and Pyke, A. (2011). Ross River Virus. In Molecular Detection of Human Viral
790 Pathogens, (Boca Raton, FL, USA: CRC Press), pp. 349–359.

791 Harley, D., Sleigh, A., and Ritchie, S. (2001). Ross River virus transmission, infection, and
792 disease: a cross-disciplinary review. Clinical Microbiology Reviews *14*, 909–932.

793 Harvey, E., Rose, K., Eden, J.-S., Lo, N., Abeyasuriya, T., Shi, M., Doggett, S.L., and Holmes, E.C.
794 (2019). Extensive diversity of RNA viruses in Australian ticks. J. Virol. *93*, e01358-18.

795 Hobson-Peters, J., Harrison, J.J., Watterson, D., Hazlewood, J.E., Vet, L.J., Newton, N.D.,
796 Warrilow, D., Colmant, A.M.G., Taylor, C., Huang, B., et al. (2019). A recombinant platform for
797 flavivirus vaccines and diagnostics using chimeras of a new insect-specific virus. Science
798 Translational Medicine *11*.

799 Illumina (2017). Effects of index misassignment on multiplexing and downstream analysis.
800 [White paper]. Retrieved November 25, 2019 from
801 https://www.illumina.com/content/dam/illumina-
802 marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-
803 004.pdf.

804 Jansen, C.C., Hemmerter, S., van den Hurk, A.F., Whelan, P.I., and Beebe, N.W. (2013).
805 Morphological versus molecular identification of *Culex annulirostris* Skuse and *Culex palpalis*
806 Taylor: key members of the *Culex sitiens* (Diptera: Culicidae) subgroup in Australasia.
807 Australian Journal of Entomology *52*, 356–362.

808 Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., and Daszak, P.
809 (2008). Global trends in emerging infectious diseases. Nature *451*, 990–993.

810 Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version
811 7: improvements in performance and usability. Molecular Biology and Evolution *30*, 772–780.

812 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper,
813 A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable
814 desktop software platform for the organization and analysis of sequence data. Bioinformatics
815 *28*, 1647–1649.

816 Knope, K., Doggett, S.L., Jansen, C.C., Johansen, C.A., Kurucz, N., Feldman, R., Lynch, S.E.,
817 Hobby, M.P., Sly, A., Jardine, A., et al. (2019). Arboviral diseases and malaria in Australia,
818 2014–15: Annual report of the National Arbovirus and Malaria Advisory Committee.
819 Communicable Diseases Intelligence (2018) *43*.

Kramer, L.D., Chiles, R.E., Do, T.D., and Fallah, H.M. (2001). Detection of St. Louis encephalitis and western equine encephalomyelitis RNA in mosquitoes tested without maintenance of a cold chain. Journal of the American Mosquito Control Association *17*, 213–215.

Larsson, A.J.M., Stanley, G., Sinha, R., Weissman, I.L., and Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. Nature Methods *15*, 305–307.

Lemmon, G.H., and Gardner, S.N. (2008). Predicting the sensitivity and specificity of published real-time PCR assays. Annals of Clinical Microbiology and Antimicrobials *7*, 18.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv:1303.3997 [q-Bio].

Li, C.-X., Shi, M., Tian, J.-H., Lin, X.-D., Kang, Y.-J., Chen, L.-J., Qin, X.-C., Xu, J., Holmes, E.C., and Zhang, Y.-Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. ELife *4*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lynch, S.E., Mee, P.T., Batovska, J., Brown, K., Crowder, J., and Rodoni, B.C. (2020). Mosquito and arbovirus biodiversity and Ross River virus genomic epidemiology in Victoria, Australia, 2016-2017. *Manuscript in preparation*.

Mackenzie, J.S., Lindsay, M.D., Coelen, R.J., Broom, A.K., Hall, R.A., and Smith, D.W. (1994). Arboviruses causing human disease in the Australasian zoogeographic region. Archives of Virology *136*, 447–467.

Mariadassou, M., Pichon, S., and Ebert, D. (2015). Microbial ecosystems are dominated by specialist taxa. Ecology Letters *18*, 974–982.

Mavale, M., Sudeep, A., Gokhale, M., Hundekar, S., Parashar, D., Ghodke, Y., Arankalle, V., and Mishra, A.C. (2012). Persistence of viral RNA in chikungunya virus-infected *Aedes aegypti* (Diptera: Culicidae) mosquitoes after prolonged storage at 28°C. The American Journal of Tropical Medicine and Hygiene *86*, 178–180.

Mayer, S.V., Tesh, R.B., and Vasilakis, N. (2017). The emergence of arthropod-borne viral diseases: a global prospective on dengue, chikungunya and Zika fevers. Acta Tropica *166*, 155–163.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE *8*, e61217.

Michie, A., Dhanasekaran, V., Lindsay, M.D.A., Neville, P.J., Nicholson, J., Jardine, A., Mackenzie, J.S., Smith, D.W., and Imrie, A. (2020). Genome-scale phylogeny and evolutionary analysis of Ross River virus reveals periodic sweeps of lineage dominance in Western Australia, 1977-2014. Journal of Virology *94*.

Murray-Darling Basin Authority (2016). River Murray weekly report for the week ending Wednesday, 5 October 2016 (Australian Government).

O'Brien, C.A., Hobson-Peters, J., Yam, A.W.Y., Colmant, A.M.G., McLean, B.J., Prow, N.A., Watterson, D., Hall-Mendelin, S., Warrilow, D., Ng, M.-L., et al. (2015). Viral RNA intermediates as targets for detection and discovery of novel and emerging mosquito-borne viruses. PLOS Neglected Tropical Diseases *9*, e0003629.

Öhlund, P., Lundén, H., and Blomström, A.-L. (2019). Insect-specific virus evolution and potential effects on vector competence. Virus Genes *55*, 127–137.

Pettersson, J.H.-O., Shi, M., Eden, J.-S., Holmes, E.C., and Hesson, J.C. (2019). Meta-transcriptomic comparison of the RNA viromes of the mosquito vectors *Culex pipiens* and *Culex torrentium* in northern Europe. Viruses *11*.

Pickering, P., Aaskov, J.G., and Liu, W. (2019). Complete genomic sequence of an Australian Sindbis virus isolated 44 years ago reveals unique indels in the E2 and nsP3 proteins. Microbiology Resource Announcements *8*.

Piper, A.M., Batovska, J., Cogan, N.O.I., Weiss, J., Cunningham, J.P., Rodoni, B.C., and Blacket, M.J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience *8*.

Pollett, S., Fauver, J.R., Maljkovic Berry, I., Melendrez, M., Morrison, A., Gillis, L.D., Johansson, M.A., Jarman, R.G., and Grubaugh, N.D. (2020). Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. The Journal of Infectious Diseases *221*, S308–S318.

Priyamvada, L., Quicke, K.M., Hudson, W.H., Onlamoon, N., Sewatanon, J., Edupuganti, S., Pattanapanyasat, K., Chokephaibulkit, K., Mulligan, M.J., Wilson, P.C., et al. (2016). Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus. PNAS *113*, 7852–7857.

Pyke, A.T., and Warrilow, D. (2016). Archival collections are important in the study of the biology, diversity, and evolution of arboviruses. Evolutionary Bioinformatics Online *12*, 27–30.

Rambaut, A. (2019). FigTree v1.4: Tree Figure Drawing Tool (Available: https://github.com/rambaut/figtree/releases Accessed 2019 Nov 20).

Ramos-Nino, M.E., Fitzpatrick, D.M., Eckstrom, K.M., Tighe, S., Hattaway, L.M., Hsueh, A.N., Stone, D.M., Dragon, J.A., and Cheetham, S. (2020). Metagenomic analysis of *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes from Grenada, West Indies. PLOS ONE *15*.

Reuter, G., Pankovics, P., Delwart, E., and Boros, Á. (2015). A novel posavirus-related single-stranded RNA virus from fish (*Cyprinus carpio*). Archives of Virology *160*, 565–568.

Ritchie, S.A., Pyke, A.T., Smith, G.A., Northill, J.A., Hall, R.A., van den Hurk, A.F., Johansen, C.A., Montgomery, B.L., and Mackenzie, J.S. (2003). Field evaluation of a sentinel mosquito

(Diptera: Culicidae) trap system to detect Japanese encephalitis in remote Australia. Journal of Medical Entomology *40*, 249–252.

Rohe, D.L., and Fall, R.P. (1979). A miniature battery powered CO2 baited light trap for mosquito-borne encephalitis surveillance. Bulletin of the Society of Vector Ecologists *4*, 24–27.

Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V., Kropinski, A.M., Krupovic, M., Kuhn, J.H., Lavigne, R., Brister, J.R., Varsani, A., et al. (2019). Minimum Information about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology *37*, 29–37.

RStudio Team (2015). RStudio: Integrated development for R. (Boston, MA).

Sadeghi, M., Altan, E., Deng, X., Barker, C.M., Fang, Y., Coffey, L.L., and Delwart, E. (2018). Virome of > 12 thousand *Culex* mosquitoes from throughout California. Virology *523*, 74–88.

Saleh, S.M., Poidinger, M., Mackenzie, J.S., Broom, A.K., Lindsay, M.D., and Hall, R.A. (2003). Complete genomic sequence of the Australian south-west genotype of Sindbis virus: comparisons with other Sindbis strains and identification of a unique deletion in the 3'-untranslated region. Virus Genes *26*, 317–327.

Schlaberg, R., Chiu, C.Y., Miller, S., Procop, G.W., Weinstock, G., Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, and Microbiology Resource Committee of the College of American Pathologists (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Archives of Pathology & Laboratory Medicine *141*, 776–786.

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR inhibitors – occurrence, properties and removal. Journal of Applied Microbiology *113*, 1014–1026.

Selck, F.W., Adalja, A.A., and Boddie, C.R. (2014). An estimate of the global health care and lost productivity costs of dengue. Vector Borne and Zoonotic Diseases (Larchmont, N.Y.) *14*, 824–826.

Shchetinin, A.M., Lvov, D.K., Deriabin, P.G., Botikov, A.G., Gitelman, A.K., Kuhn, J.H., and Alkhovsky, S.V. (2015). Genetic and phylogenetic characterization of Tataguine and Witwatersrand viruses and other orthobunyaviruses of the Anopheles A, Capim, Guamá, Koongol, Mapputta, Tete, and Turlock serogroups. Viruses *7*, 5987–6008.

Shi, M., Neville, P., Nicholson, J., Eden, J.-S., Imrie, A., and Holmes, E.C. (2017). High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. Journal of Virology *91*, e00680-17.

Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. The Journal of General Virology *96*, 1193–1206.

Sozhamannan, S., Holland, M.Y., Hall, A.T., Negrón, D.A., Ivancich, M., Koehler, J.W., Minogue, T.D., Campbell, C.E., Berger, W.J., Christopher, G.W., et al. (2015). Evaluation of signature

930 erosion in Ebola virus due to genomic drift and its impact on the performance of diagnostic
931 assays. Viruses *7*, 3130–3154.

932 Steinegger, M., and Salzberg, S.L. (2020). Terminating contamination: large-scale search
933 identifies more than 2,000,000 contaminated entries in GenBank. Genome Biology *21*, 115.

934 Stellrecht, K.A. (2018). The drift in molecular testing for influenza: mutations affecting assay
935 performance. Journal of Clinical Microbiology *56*.

936 Sutherland, G.L., and Nasci, R.S. (2007). Detection of West Nile virus in large pools of
937 mosquitoes. Journal of the American Mosquito Control Association *23*, 389–395.

938 Tangudu, C.S., Charles, J., Hurt, S.L., Dunphy, B.M., Smith, R.C., Bartholomay, L.C., and Blitvich,
939 B.J. (2019). Skunk River virus, a novel orbivirus isolated from *Aedes trivittatus* in the United
940 States. The Journal of General Virology *100*, 295–300.

941 Tay, W.T., Kerr, P.J., and Jermiin, L.S. (2016). Population genetic structure and potential
942 incursion pathways of the bluetongue virus vector *Culicoides brevitarsis* (Diptera:
943 Ceratopogonidae) in Australia. PLOS ONE *11*, e0146699.

944 Vasilakis, N., Tesh, R.B., Popov, V.L., Widen, S.G., Wood, T.G., Forrester, N.L., Gonzalez, J.P.,
945 Saluzzo, J.F., Alkhovsky, S., Lam, S.K., et al. (2019). Exploiting the legacy of the arbovirus
946 hunters. Viruses *11*.

947 Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (New York: Springer-Verlag).

948 Wilder-Smith, A., Gubler, D.J., Weaver, S.C., Monath, T.P., Heymann, D.L., and Scott, T.W.
949 (2017). Epidemic arboviral diseases: priorities for research and public health. The Lancet
950 Infectious Diseases *17*, e101–e106.

951 Wille, M., Shi, M., Klaassen, M., Hurt, A.C., and Holmes, E.C. (2019). Virome heterogeneity
952 and connectivity in waterfowl and shorebird communities. The ISME Journal *13*, 2603–2616.

953 Zakrzewski, M., Rašić, G., Darbro, J., Krause, L., Poo, Y.S., Filipović, I., Parry, R., Asgari, S.,
954 Devine, G., and Suhrbier, A. (2018). Mapping the virome in wild-caught *Aedes aegypti* from
955 Cairns and Bangkok. Scientific Reports *8*, 1–12.

956

Tables and Figures

Table 1: Mosquito counts for the 21 traps used in this study, sampled over seven weeks (45 – 51) in 2016 in Gannawarra, Mildura and Wellington. The traps were first subsampled for cell culture-based arbovirus screening by the Victorian Arbovirus Disease Control Program (VADCP), and then the rest of the trap was split into differently sized subsamples for metatranscriptomic sequencing. The sequencing subsample sizes consisted of 100, 1,000 and the remaining mosquitoes in each trap. The mosquito counts are based on weight.

Location	**Week**	**Estimated total mosquito count**	**Mosquitoes screened via cell culture**	**Mosquitoes screened via sequencing**	**Number of sequencing subsamples**
Gannawarra	45	4,595	370	4,225	6
	46	2,502	152	2,350	4
	47	1,992	148	1,844	3
	48	1,652	182	1,470	3
	49	1,157	351	806	2
	50	2,154	347	1,807	3
	51	3,814	358	3,456	5
Mildura	45	618	340	278	2
	46	2,014	149	1,865	3
	47	2,051	152	1,899	3
	48	3,958	349	3,609	5
	49	2,356	350	2,006	3
	50	1,189	348	841	2
	51	1,706	350	1,356	3
Wellington	45	4,829	349	4,480	6
	46	1,535	152	1,383	3
	47	13,659	264	13,395	15
	48	1,497	222	1,275	3
	49	5,226	336	4,890	6
	50	1,186	366	820	2
	51	2,528	350	2,178	4
Total:		**62,218**	**5,985**	**56,233**	**86**

965 **Table 2:** Recommendations for the incorporation of metatranscriptomics in routine arbovirus
966 surveillance programs.

Category	Recommendation	Outcomes
Sampling	Sequence all the mosquitoes collected from surveillance traps.	Increased likelihood of detecting arboviral activity.
	Investigate other sample types for metatranscriptomic sequencing (e.g. gravid mosquitoes, FTA cards)	Less sequencing of uninfected mosquitoes, thereby improving arbovirus detection sensitivity.
Laboratory protocol	Develop an in-house ribosomal RNA depletion protocol.	Greater ability to customise depletion, thereby improving arbovirus detection sensitivity; decreased assay cost.
	Use unique dual indexing to multiplex samples.	Reduced index cross-talk, thereby improving arbovirus detection sensitivity.
	Automate library preparation where possible.	Increased multiplexing capacity; decreased assay cost and turnaround time.
	Use ultra-high-throughput sequencing (e.g. Illumina NovaSeq).	Increased multiplexing capacity; decreased assay cost; faster turnaround than HiSeq.
Bioinformatics	Establish a user-friendly bioinformatics pipeline.	Less reliance on specialised bioinformatics expertise.
	Develop a tool to assess the risk of novel or unclassified viruses.	Faster, less complex data analysis with relevant reporting for public health.
	Formulate an organised and cost-effective storage plan for high volumes of sequencing data.	Ability to repurpose or re-analyse past metatranscriptomic surveillance data with updated databases and bioinformatics.
Reference databases	Establish a DNA barcode database of local mosquito species.	Comprehensive identification of mosquito species in surveillance traps.
	Acquire whole genome sequences of arbovirus isolates for inclusion in databases used for screening data.	Improved arbovirus detection sensitivity; high-resolution phylogenetics to determine arbovirus origins and dispersal.
	Curate a contamination database by sequencing samples from laboratory surfaces and reagents.	Improved ability to distinguish real signal from background or contamination.
Quality control	Include negative and positive controls.	Detection of contamination; ability to assess assay validity.
	Standardise laboratory and bioinformatics protocols.	Consistent, reproducible surveillance results of known sensitivity and specificity.
	Regularly validate assay sensitivity and specificity in response to protocol modifications.	Enables protocol updates while ensuring adequate assay sensitivity and appropriate detection thresholds.
	Confirm important arbovirus detections with RT-qPCR.	Confidence in reliability of arbovirus detections for public health reporting.

967

Figure 1: Map of the three locations traps were set up in Victoria, Australia. The highlighted regions represent the local government areas of Mildura, Gannawarra, and Wellington.

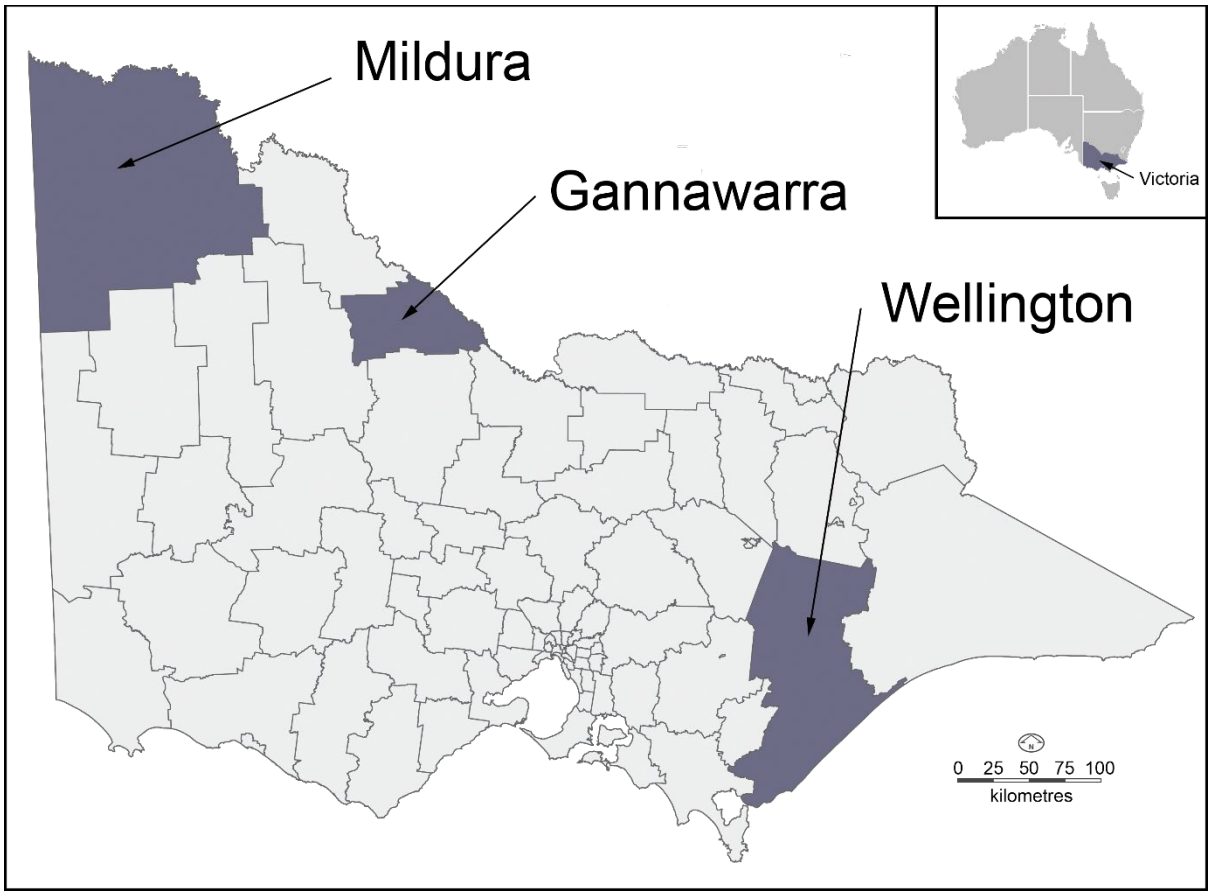


Figure 2: The estimated total number of mosquitoes used for sequencing per location per week from 7 November 2016 (Week 45) to 19 December 2016 (Week 51).

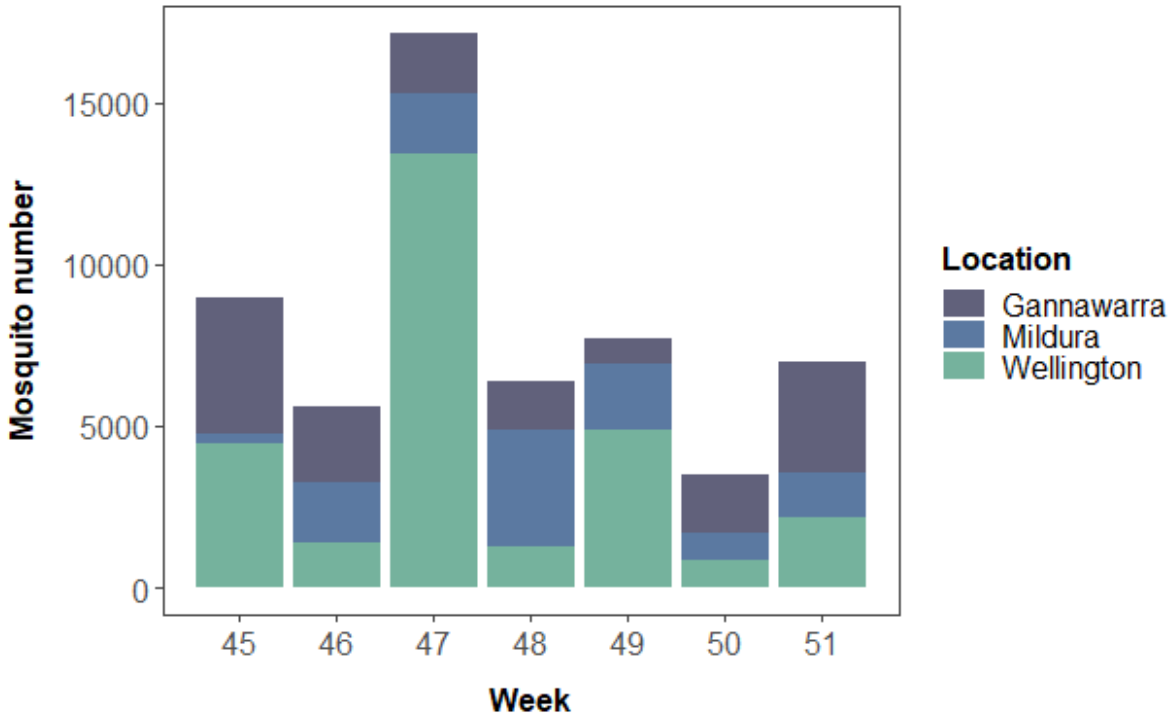


Figure 3: Domain-level taxonomic composition of the individual mosquito trap samples and control samples. Abundance is based on read counts for taxonomically classified contigs produced by *de novo* assembly.

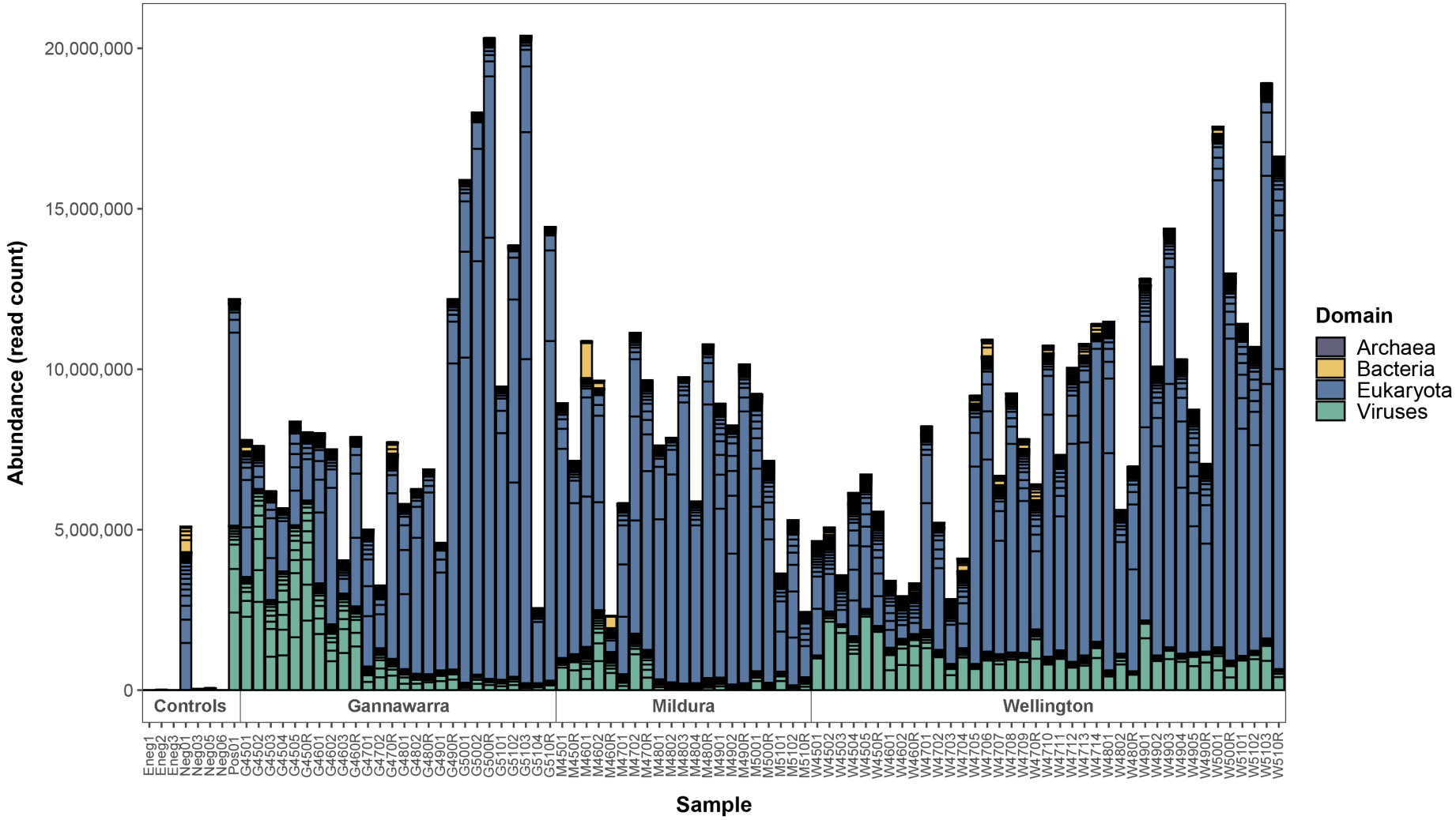


Figure 4: Mosquito and biting midge species identification based on metatranscriptomic sequencing (MSeq) of traps from Gannawarra, Mildura and Wellington over seven weeks (45 – 51) in 2016. Relative abundance estimations derived from a COI-based analysis is shown in (A) when using all subsamples, and (B) when using only the 100-mosquito subsamples. Members of the *Culex pipiens* species complex that cannot be differentiated by COI have been conglomerated.

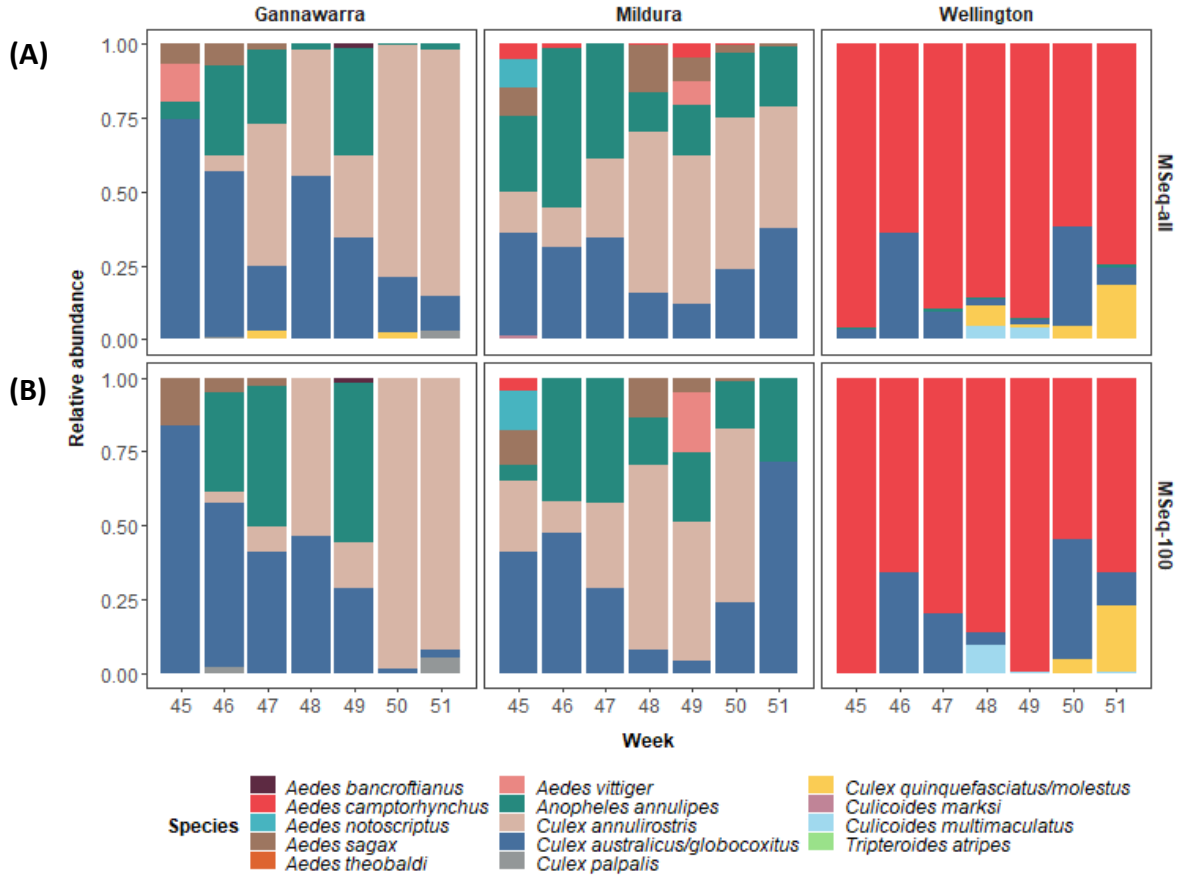


Figure 5: Arbovirus detections based on metatranscriptomic sequencing (MSeq) of traps from Gannawarra, Mildura and Wellington over seven weeks (45 – 51) in 2016. Positive detection was based on a Percent Coverage by Reads ratio (%CR-r) ≥ 2, which is calculated by taking the percent coverage of the arbovirus genome by trap reads and dividing it by the percent coverage by negative control reads. Detections are shown in (A) when using all subsamples, and (B) when using only the 100-mosquito subsamples. The size of the dots is proportional to the number of reads contributing to each arbovirus detection.

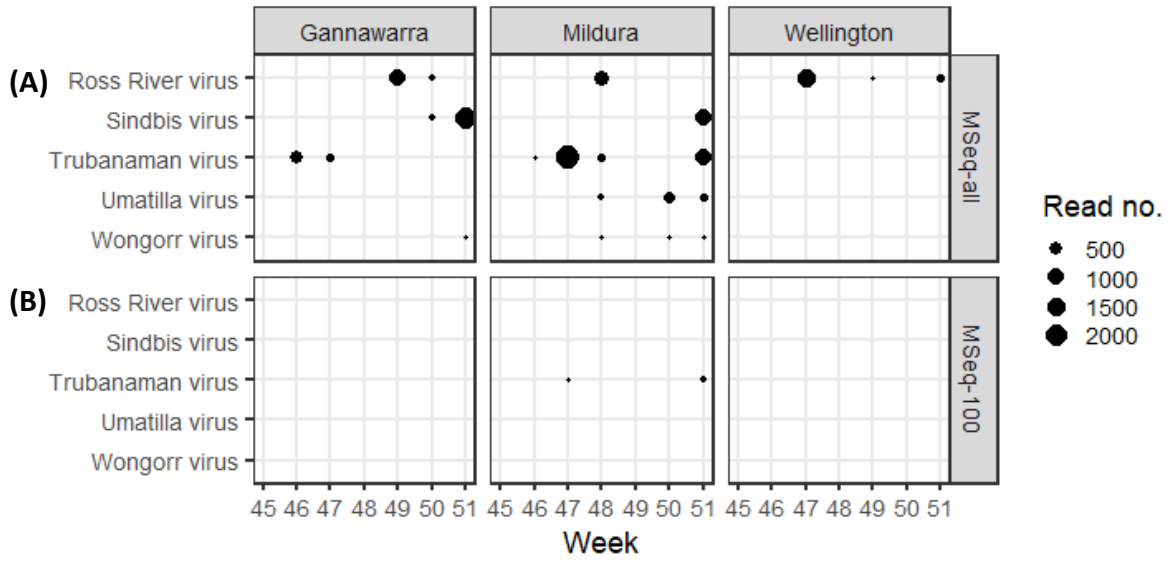


Figure 6: Surveillance data produced by the Victorian Arbovirus Disease Control Program (VADCP) for the initial subsamples taken from the traps in Gannawarra, Mildura and Wellington over seven weeks (45 – 51) in 2016 (Table 1). (A) The relative abundance of the mosquito species in the initial trap subsamples, based on morphological identification. It should be noted that the VADCP does not survey for biting midge species. (B) Arboviruses detected in the subsampled mosquitoes from each trap, based on cell culture screening.

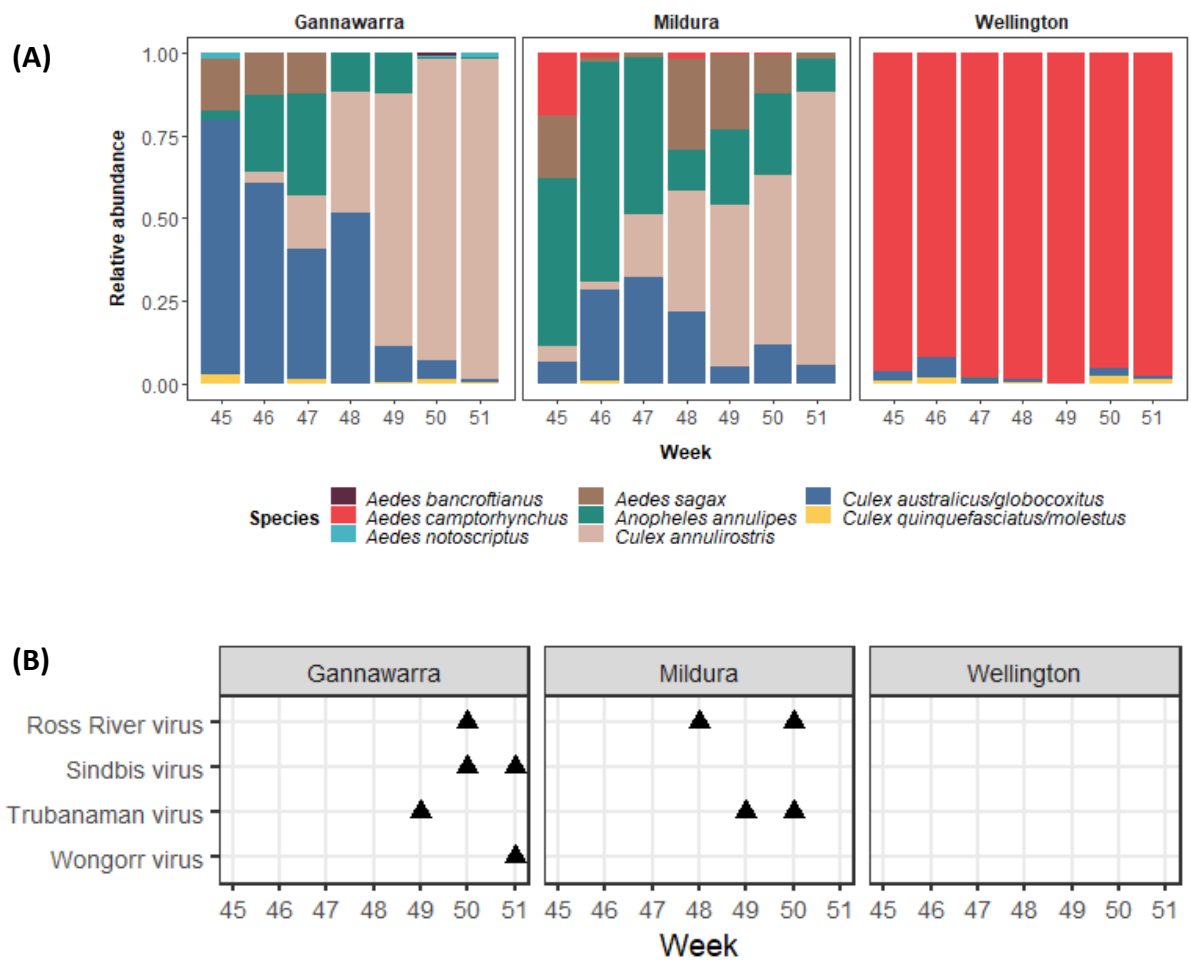
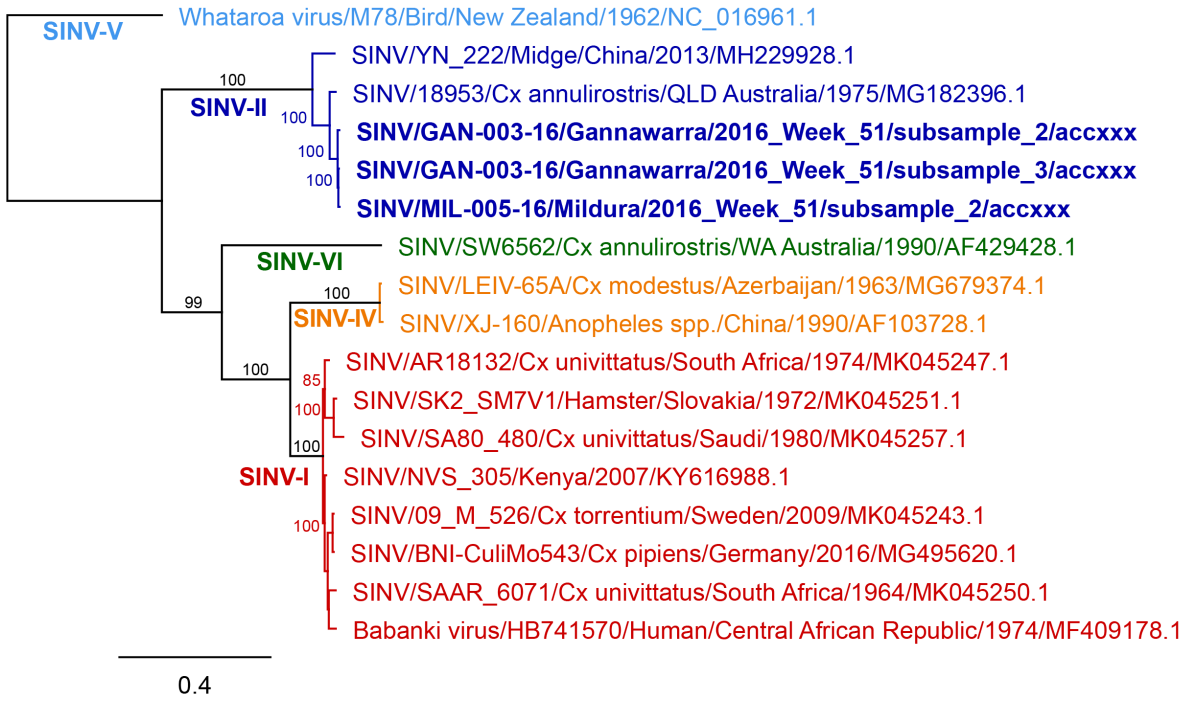
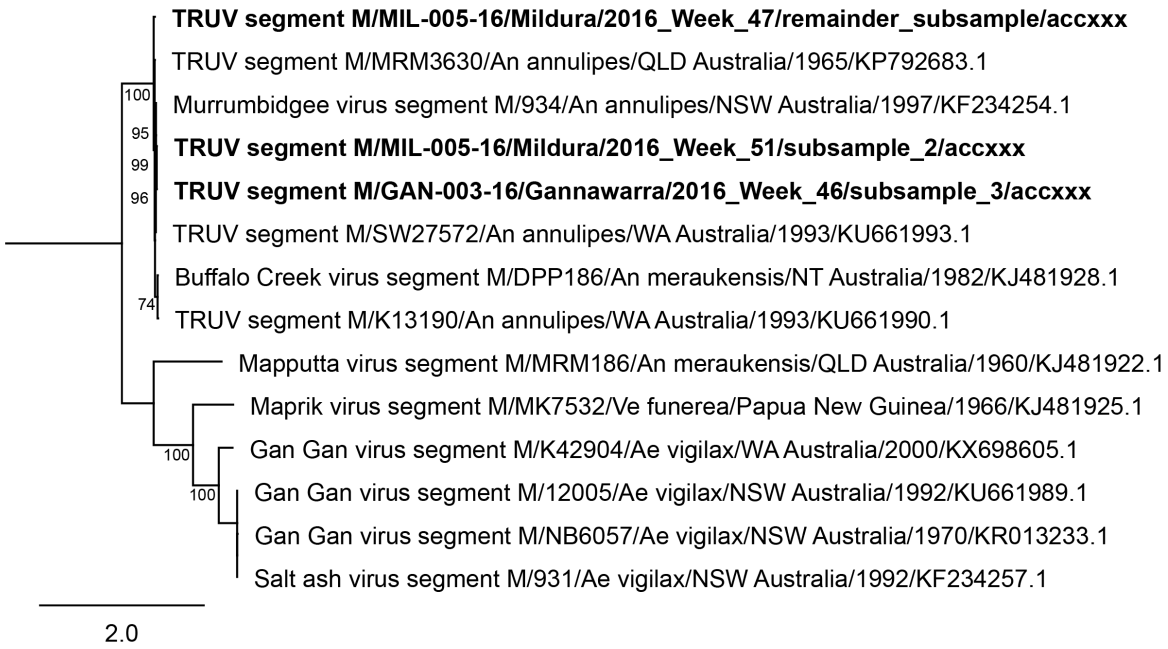


Figure 7: Phylogenetic trees depicting the genetic relationship of the (A) Ross River virus (RRV), (B) Sindbis virus (SINV), (C) Trubanaman virus (TRUV) and (D) Umatilla virus (UMAV) detections. The trees are based on whole genome sequence alignments for RRV (11,362 nt) and SINV (11,460 nt), Segment M for TRUV (4,152 nt), and Segment 7 for UMAV (1,364 nt). A maximum likelihood (GTR model) analysis was used with 1,000 bootstrap replicates (only values >70% shown). Coloured clades represent virus genotypes, as indicated by the clade label. The naming convention for other viruses is virus/strain/host/location/year/GenBank accession, with boldface indicating sequences generated in this study.

(B)


SINV-V
Whataroa virus/M78/Bird/New Zealand/1962/NC_016961.1
SINV/YN_222/Midge/China/2013/MH229928.1
SINV-II
SINV/18953/Cx annulirostris/QLD Australia/1975/MG182396.1
SINV/GAN-003-16/Gannawarra/2016_Week_51/subsample_2/accxxx
SINV/GAN-003-16/Gannawarra/2016_Week_51/subsample_3/accxxx
SINV/MIL-005-16/Mildura/2016_Week_51/subsample_2/accxxx
SINV-VI
SINV/SW6562/Cx annulirostris/WA Australia/1990/AF429428.1
SINV-IV
SINV/LEIV-65A/Cx modestus/Azerbaijan/1963/MG679374.1
SINV/XJ-160/Anopheles spp./China/1990/AF103728.1
SINV/AR18132/Cx univittatus/South Africa/1974/MK045247.1
SINV/SK2_SM7V1/Hamster/Slovakia/1972/MK045251.1
SINV/SA80_480/Cx univittatus/Saudi/1980/MK045257.1
SINV-I
SINV/NVS_305/Kenya/2007/KY616988.1
SINV/09_M_526/Cx torrentium/Sweden/2009/MK045243.1
SINV/BNI-CuliMo543/Cx pipiens/Germany/2016/MG495620.1
SINV/SAAR_6071/Cx univittatus/South Africa/1964/MK045250.1
Babanki virus/HB741570/Human/Central African Republic/1974/MF409178.1
100
100
100
100
99
100
100
85
100
100
100
0.4

(C)


TRUV segment M/MIL-005-16/Mildura/2016_Week_47/remainder_subsample/accxxx
TRUV segment M/MRM3630/An annulipes/QLD Australia/1965/KP792683.1
Murrumbidgee virus segment M/934/An annulipes/NSW Australia/1997/KF234254.1
TRUV segment M/MIL-005-16/Mildura/2016_Week_51/subsample_2/accxxx
TRUV segment M/GAN-003-16/Gannawarra/2016_Week_46/subsample_3/accxxx
TRUV segment M/SW27572/An annulipes/WA Australia/1993/KU661993.1
Buffalo Creek virus segment M/DPP186/An meraukensis/NT Australia/1982/KJ481928.1
TRUV segment M/K13190/An annulipes/WA Australia/1993/KU661990.1
Mapputta virus segment M/MRM186/An meraukensis/QLD Australia/1960/KJ481922.1
Maprik virus segment M/MK7532/Ve funerea/Papua New Guinea/1966/KJ481925.1
Gan Gan virus segment M/K42904/Ae vigilax/WA Australia/2000/KX698605.1
Gan Gan virus segment M/12005/Ae vigilax/NSW Australia/1992/KU661989.1
Gan Gan virus segment M/NB6057/Ae vigilax/NSW Australia/1970/KR013233.1
Salt ash virus segment M/931/Ae vigilax/NSW Australia/1992/KF234257.1
100
95
99
96
74
100
100
2.0

(D)

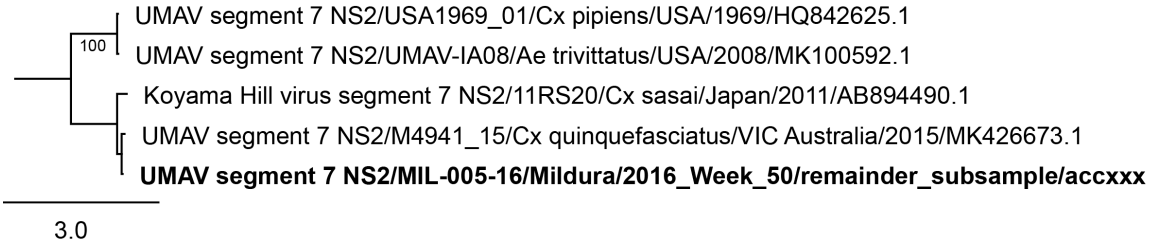

UMAV segment 7 NS2/USA1969_01/Cx pipiens/USA/1969/HQ842625.1
UMAV segment 7 NS2/UMAV-IA08/Ae trivittatus/USA/2008/MK100592.1
Koyama Hill virus segment 7 NS2/11RS20/Cx sasai/Japan/2011/AB894490.1
UMAV segment 7 NS2/M4941_15/Cx quinquefasciatus/VIC Australia/2015/MK426673.1
UMAV segment 7 NS2/MIL-005-16/Mildura/2016_Week_50/remainder_subsample/accxxx
100
3.0

Figure 8: Results of the virome analysis showing the BLASTx match for contigs assembled from traps in Gannawarra, Mildura and Wellington over seven weeks (45 – 51) in 2016. The dot colour indicates the similarity of the contigs to the virus GenBank sequence, as measured by percentage of identical amino acids. The dot size indicates virus abundance, which is based on the number of reads mapping to the matching contigs. Coloured names indicate highly similar or identical viruses (Figure S3) that likely represent a single detection. The arboviruses detected using the targeted database (Figure 5) are not included here.

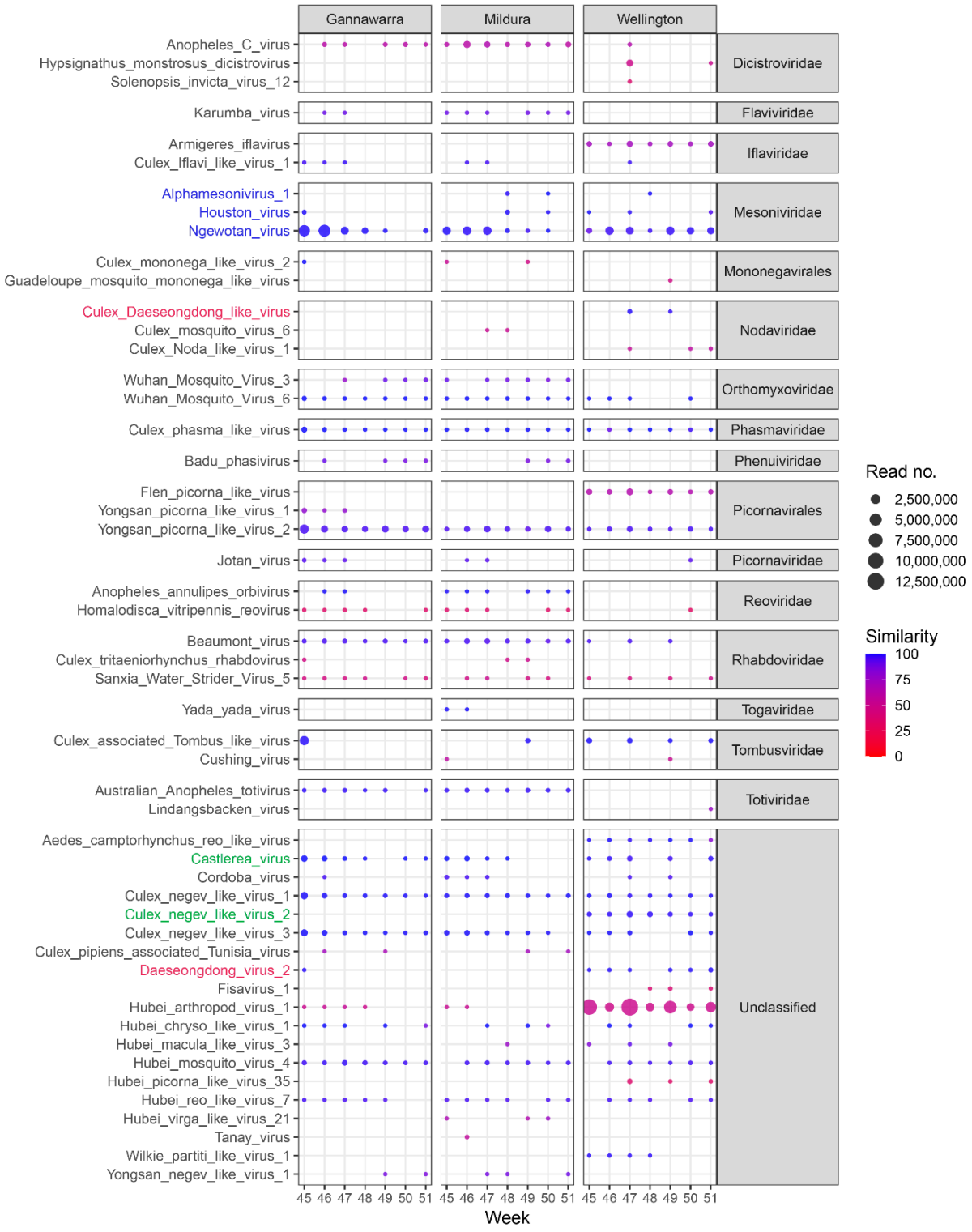
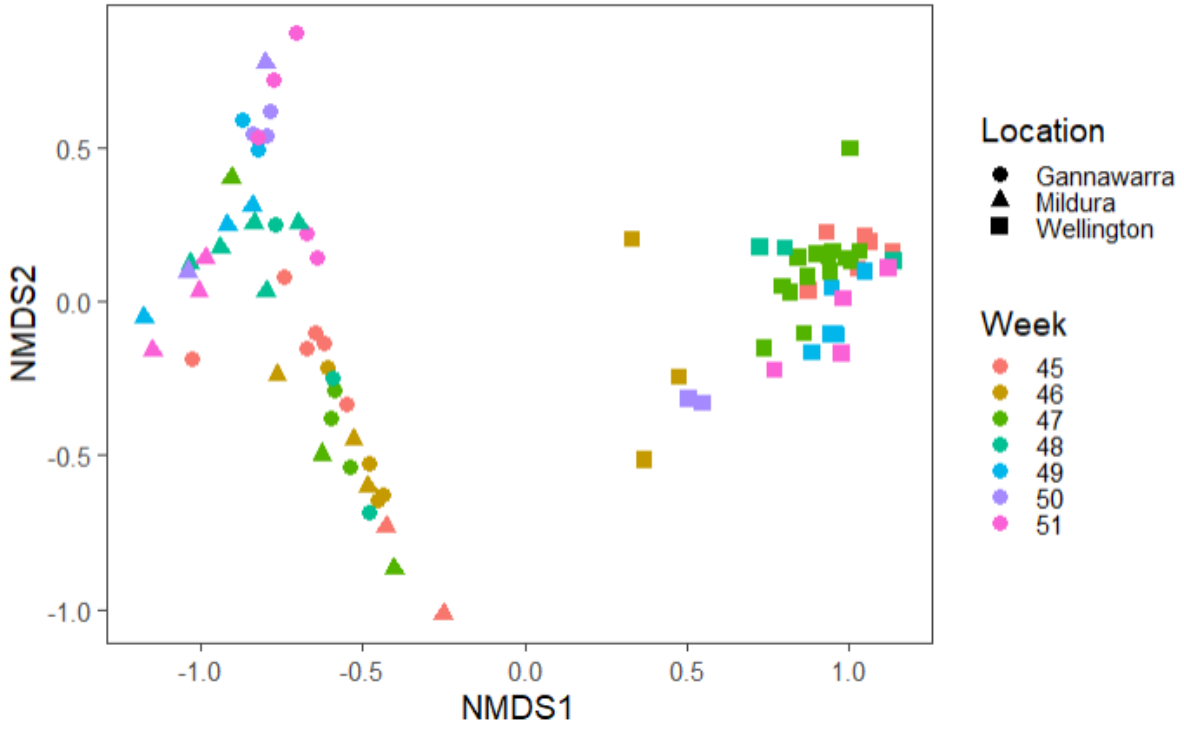
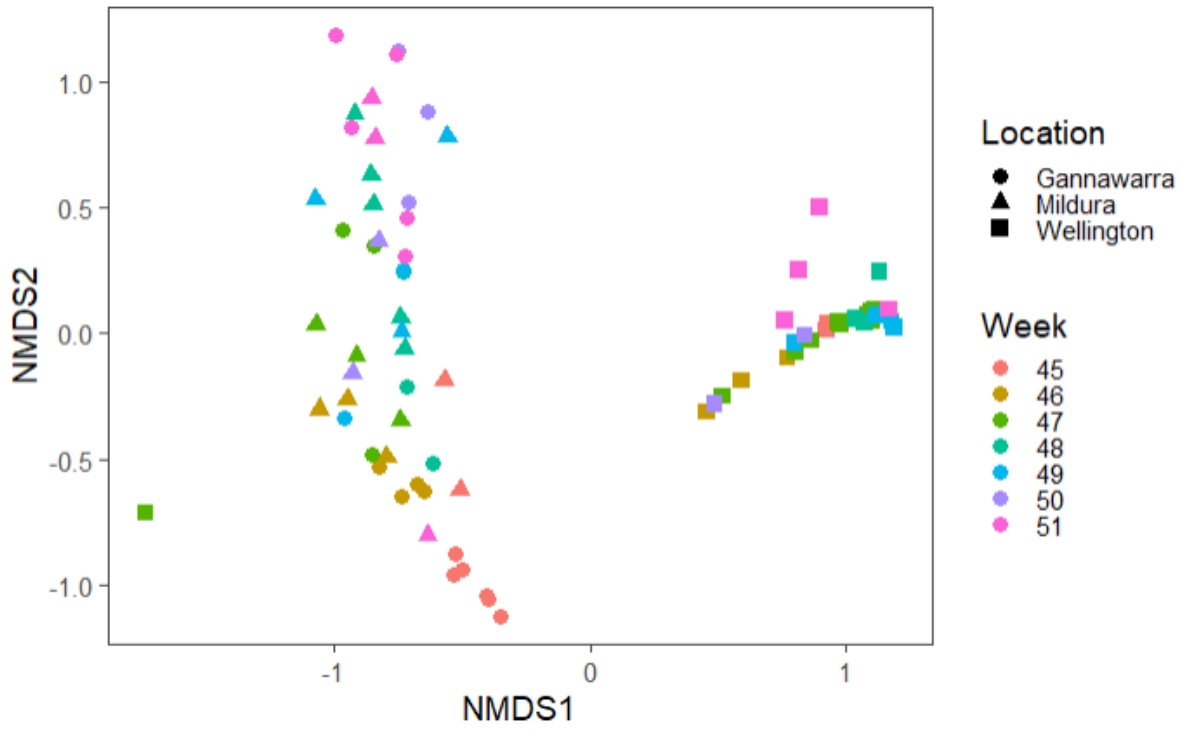


Figure 9: Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity of the (A) viruses and (B) mosquitoes in the trap subsamples from three locations over seven weeks showing a separation between coastal (Wellington) and inland (Gannawarra and Mildura) samples.

(A)



(B)



3.5 Supplementary information

Table S1: The number of 3 mm glass beads (Sigma-Aldrich) and volume of Buffer AVL (Qiagen) that was added to each sample based on the number of mosquitoes present.

Mosquito no.	Bead no.	Buffer AVL (mL)
1-100	10	1
101-200	13	2
201-300	16	3
301-400	19	4
401-500	22	5
501-600	25	6
601-700	28	7
701-800	31	8
801-900	34	9
901-1000	37	10

Table S2: Details of the reverse transcription quantitative PCR (RT-qPCR) assays used to screen trap subsamples for the following arboviruses: Ross River virus (RRV), Trubanaman virus (TRUV), Umatilla virus (UMAV), Sindbis virus (SINV) and Wongorr virus (WGRV).

Virus	Primer/probe name	Primer/probe sequence (5' → 3')	Primer conc. (µM)	Target	Amplicon size (bp)	Reference	PCR kit	PCR cycling conditions	RNA (µL)	Additional reagents
RRV	RRVE2F	ACGGAAGAAGGGATTGAGTACCA	7.5	E2	67	(Hall et al., 2011)	AgPath-ID One-Step RT-PCR Kit (Applied Biosystems)	30' 48°C; 10' 95°C; 40 x (15" 95°C, 1' 60°C)	5	N/A
	RRVE2R	TCGTCAGTTGCGCCCATA	7.5							
	RRVE2Prob	CAACAACCCGCCGGTCCGC	5							
TRUV	MUR-BUC-TRUV S -F	TGGAGACCTGGAAGTTGTTATTCA	10	N	487	(Lynch et al., 2020)		30' 48°C; 10' 95°C; 40 x (15" 95°C, 30" 56°C, 45" 60°C); 15" 95°C; 1' 56°C; 15" 95°C		10X SYBR Green I (Invitrogen)
	MUR-BUC-TRUV S 718 -R	ATTGTCTTGCAGCTTCTGACATG								
UMAV	SLOV S2 602F	TGAACCGGCCGATACAGAAT		T2	251	(Cowled et al., 2009)		30' 48°C; 10' 95°C; 40 x (15" 95°C, 1' 60°C); 15" 95°C; 1' 60°C; 15" 95°C		
	SLOV S2 852R	TGAGGGATTTGGTGGTAATGTG								
SINV	VIR966-F	TCCATGCTAATGCTAGAGCGTTTTCGCA		nsP1	98	(Eshoo et al., 2007)	Power SYBR Green RNA-to-Ct 1-Step Kit (Applied Biosystems)	30' 48°C; 10' 95°C; 40 x (15" 95°C, 1' 60°C); 15" 95°C; 15" 60°C; 15" 95°C	2	N/A
	VIR966-R	TGGCGCACTTCCAATGTCCAGGAT								
WGRV	FC54	TAATGGGTGGCAGTGCG		VP3	214	In-house	SensiFAST SYBR Lo-ROX One-Step Kit (Bioline)	15' 45°C; 2' 95°C; 40 x (5" 95°C, 10" 58°C); 10" 72°C; 15" 95°C; 1' 58°C; 15" 95°C	4	
	RC2 53	GCTGTGACATGAGGTTCATGTAATT								

Table S3: Arbovirus detections made by metatranscriptomic sequencing in the 21 mosquito traps from Gannawarra, Mildura and Wellington over seven weeks (45 – 51) in 2016 using three different positive detection criteria: (A) Fold Coverage by Reads ratio (FCR-r), resulting in 9 detections; (B) Percent Coverage by Contigs ratio (%CC-r), resulting in 15 detections; (C) Percent Coverage by Reads ratio (%CR-r), resulting in 22 detections. Ratios ≥ 2 were considered a positive detection.

(A)

FCR-r	Gannawarra							Mildura							Wellington						
	45	46	47	48	49	50	51	45	46	47	48	49	50	51	45	46	47	48	49	50	51
Ross River virus																					
Sindbis virus																					
Trubanaman virus																					
Umatilla virus																					
Wongorr virus																					

(B)

%CC-r	Gannawarra							Mildura							Wellington						
	45	46	47	48	49	50	51	45	46	47	48	49	50	51	45	46	47	48	49	50	51
Ross River virus																					
Sindbis virus																					
Trubanaman virus																					
Umatilla virus																					
Wongorr virus																					

(C)

%CR-r	Gannawarra							Mildura							Wellington						
	45	46	47	48	49	50	51	45	46	47	48	49	50	51	45	46	47	48	49	50	51
Ross River virus																					
Sindbis virus																					
Trubanaman virus																					
Umatilla virus																					
Wongorr virus																					

Table S4: Arbovirus detections in mosquito subsamples based on reverse transcription quantitative PCR compared to metatranscriptomic sequencing (MSeq). Only Ct values are reported for the Ross River virus (RRV) assay as it is probe-based, whereas both Ct and Tm values are reported for the Sindbis virus (SINV), Trubanaman virus (TRUV), Umatilla virus (UMAV), and Wongorr virus (WGRV) assays as they are SYBR-based assays. The MSeq Percent Coverage by Reads (%CR) is provided for each detection.

Virus	Location	Week	Subsample	Mosq. no.	Ct	Tm	MSeq positive	%CR
RRV	Gannawarra	49	1	100	39.99	-	No	0.00
RRV	Gannawarra	49	R	706	27.97	-	Yes	99.43
RRV	Gannawarra	50	2	1000	32.56	-	Yes	29.61
RRV	Gannawarra	50	R	707	37.74	-	No	0.00
RRV	Mildura	48	2	1000	39.82	-	No	1.73
RRV	Mildura	48	4	1000	27.8	-	Yes	98.22
RRV	Mildura	49	2	1000	39.87	-	No	0.00
RRV	Wellington	47	2	1000	28.87	-	Yes	98.28
RRV	Wellington	47	6	1000	38.56	-	Yes	4.32
RRV	Wellington	47	11	1000	39.39	-	Yes	2.15
RRV	Wellington	49	2	1000	35.83	-	Yes	3.83
RRV	Wellington	51	2	1000	34.3	-	Yes	45.22
SINV	Gannawarra	50	2	1000	*Negative*		Yes	10.06
SINV	Gannawarra	50	R	707	39.84	82	Yes	9.17
SINV	Gannawarra	51	2	1000	38.78	82	Yes	98.86
SINV	Gannawarra	51	3	1000	39.21	82	Yes	98.30
SINV	Gannawarra	51	4	1000	*Negative*		Yes	9.20
SINV	Mildura	51	2	1000	39.91	82	Yes	98.83
TRUV	Gannawarra	46	1	100	33.66	83	No	1.53
TRUV	Gannawarra	46	2	1000	27.49	83.5	Yes	28.73
TRUV	Gannawarra	46	3	1000	23.56	83	Yes	86.02
TRUV	Gannawarra	46	R	250	28.92	83	Yes	3.77
TRUV	Gannawarra	47	R	744	26.52	83.5	Yes	49.92
TRUV	Gannawarra	51	2	1000	31.2	83	No	0.00
TRUV	Mildura	46	2	1000	30.78	83	No	0.00
TRUV	Mildura	46	R	765	28.89	83.5	Yes	2.35
TRUV	Mildura	47	1	100	27.69	83	Yes	3.42
TRUV	Mildura	47	2	1000	24.58	83.5	Yes	32.12
TRUV	Mildura	47	R	799	19.07	83.5	Yes	98.43
TRUV	Mildura	48	2	1000	26.68	83.5	Yes	3.99
TRUV	Mildura	48	R	509	25.11	83.5	Yes	34.60
TRUV	Mildura	49	2	1000	30.29	83	No	1.52
TRUV	Mildura	51	1	100	26.79	83.5	Yes	14.87
TRUV	Mildura	51	2	1000	20.82	83.5	Yes	93.36
TRUV	Mildura	51	R	256	29.96	83	No	0.00
UMAV	Mildura	48	1	100	33.55	82	No	0.00
UMAV	Mildura	48	2	1000	31.03	82	Yes	10.72

UMAV	Mildura	50	R	741	26.04	81.5	Yes	55.89
UMAV	Mildura	51	2	1000	32.78	82	Yes	28.51
WGRV	Gannawarra	51	2	1000	34.7	86.5	Yes	30.34
WGRV	Mildura	48	2	1000	33.47	86.5	Yes	17.11
WGRV	Mildura	49	1	100	33.36	86	No	0.00
WGRV	Mildura	49	2	1000	33.98	86	No	0.00
WGRV	Mildura	50	R	741	33.73	85.5	Yes	7.75
WGRV	Mildura	51	2	1000	33.72	87	Yes	9.58

1047 R = remainder subsample

1048

Figure S1: Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity of the contigs in trap samples (crosses), the positive control sample (square), library negative control samples (triangles), and extraction negative control samples (circles).

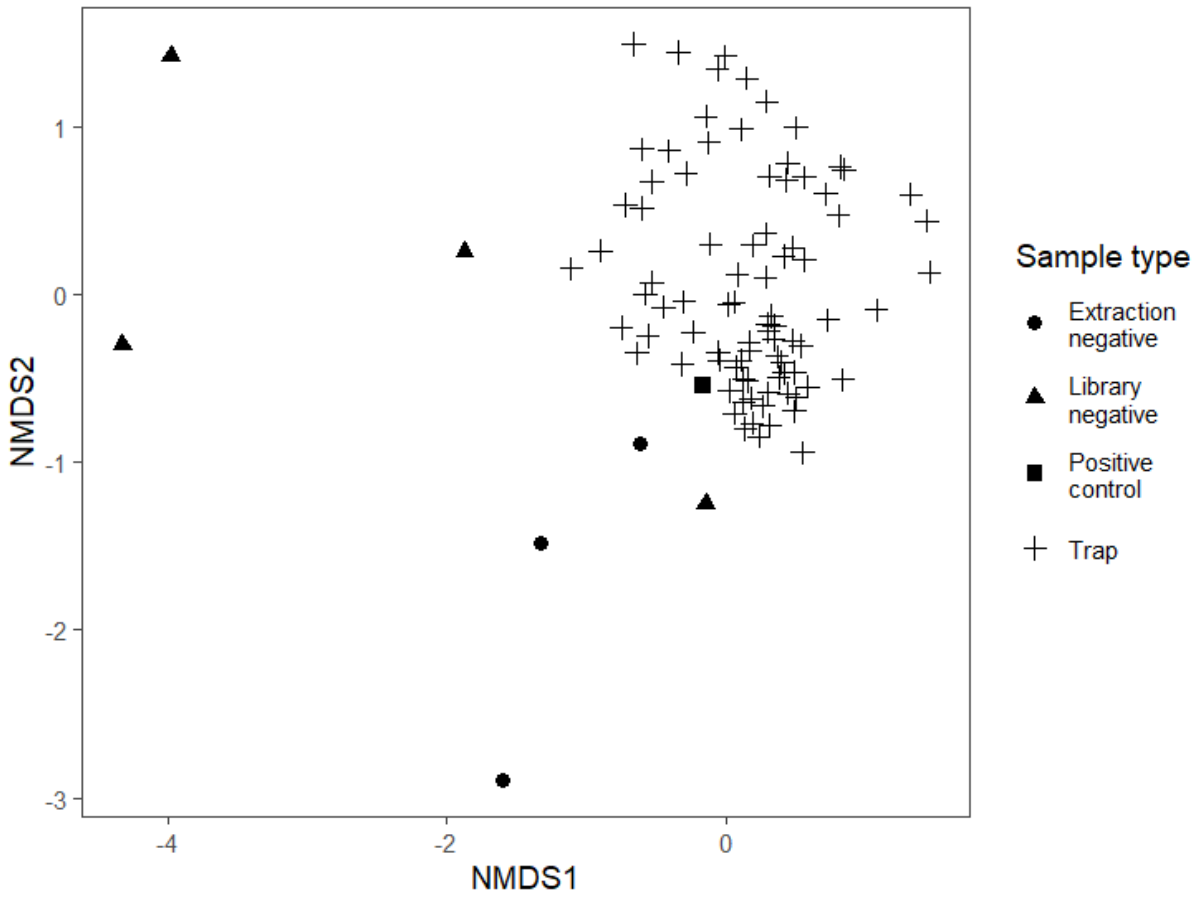


Figure S2: The relationship between Ross River virus (RRV) Ct values produced by reverse transcription quantitative PCR and RRV Percent Coverage by Reads (%CR) produced by metatranscriptomic sequencing.

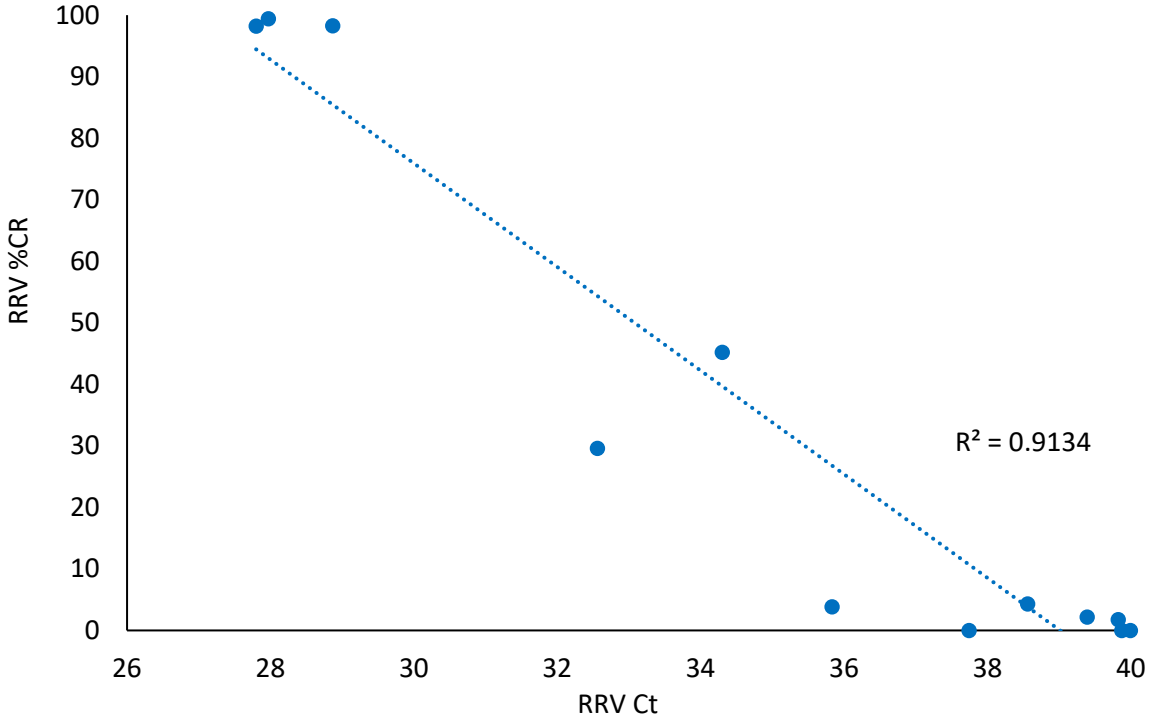


Figure S3: A maximum likelihood tree of reference GenBank RNA-dependent RNA polymerase (RdRp) sequences of the 51 viruses that were a BLASTx match for trap contigs as part of the broader virome analysis. Highly similar or identical reference RdRp sequences have been highlighted in distinct colours and indicate possible redundancies in virus detection.

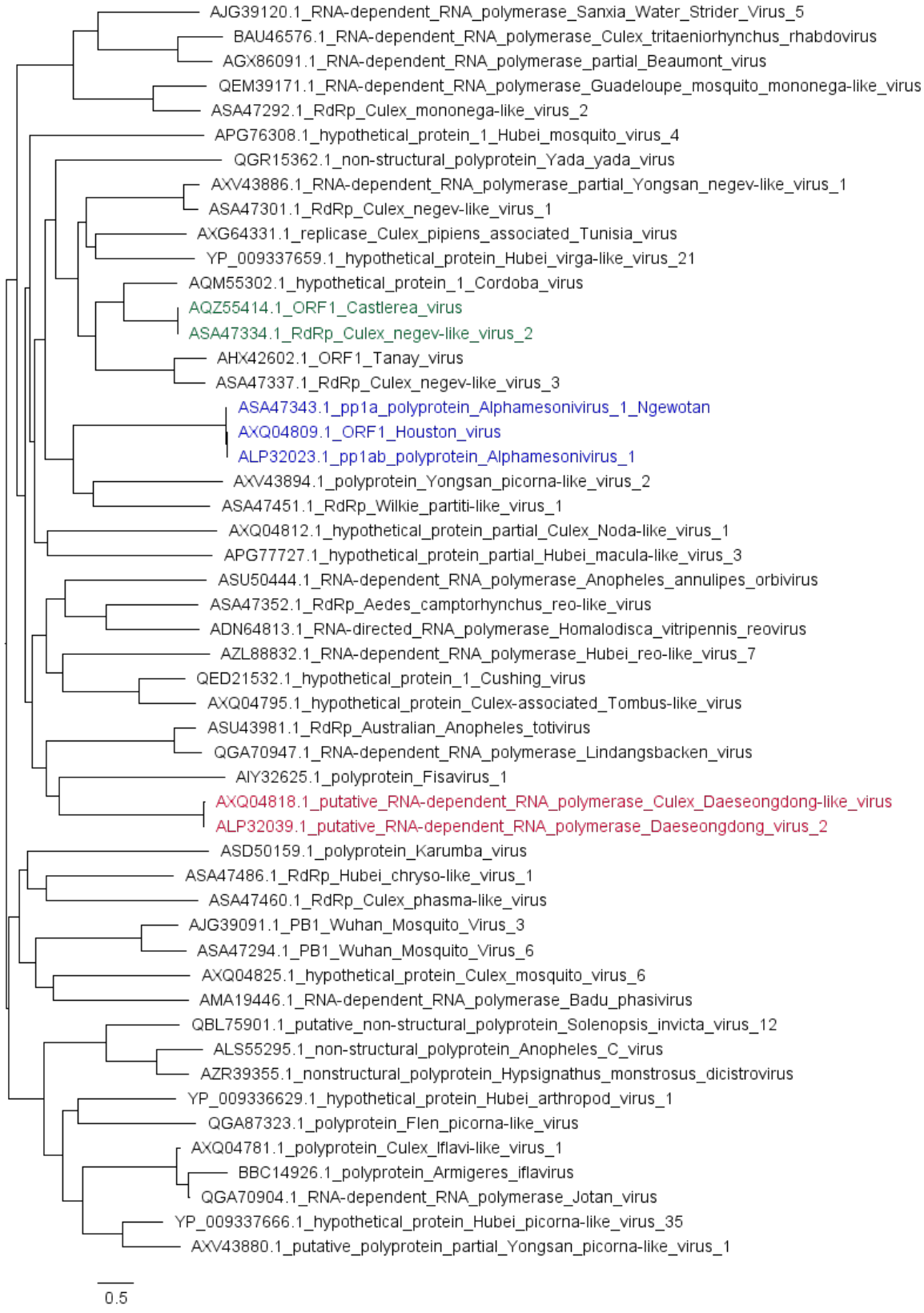
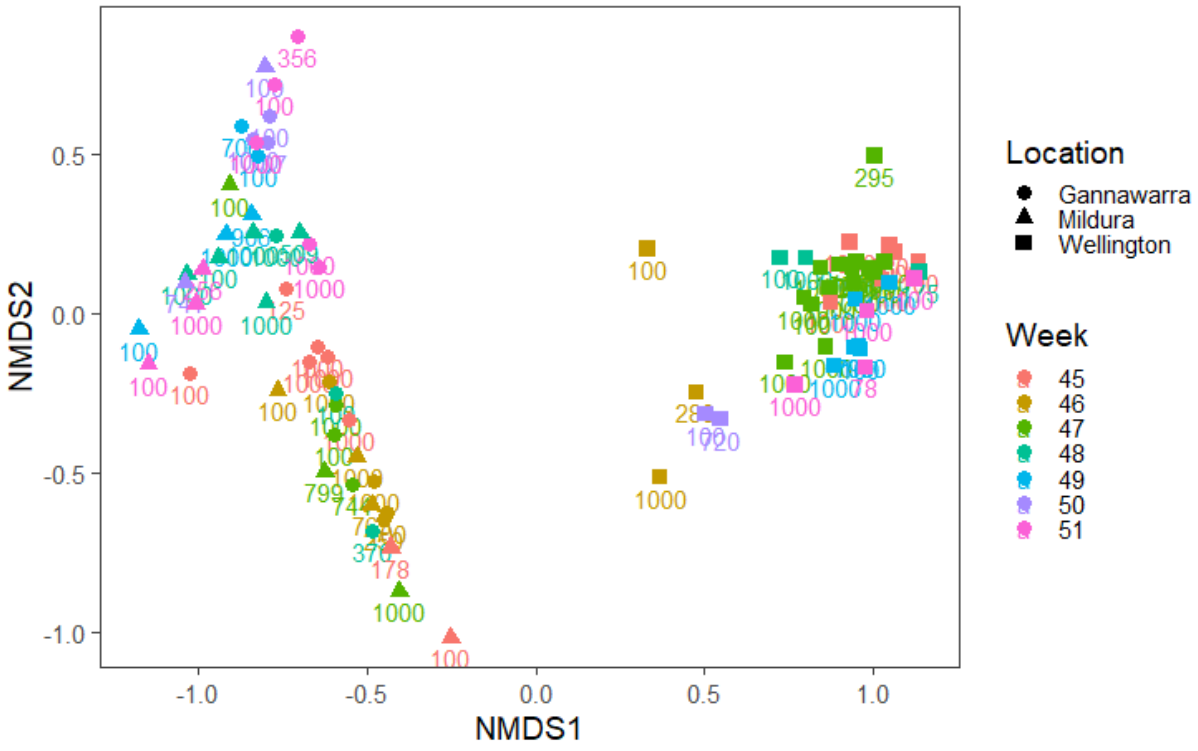
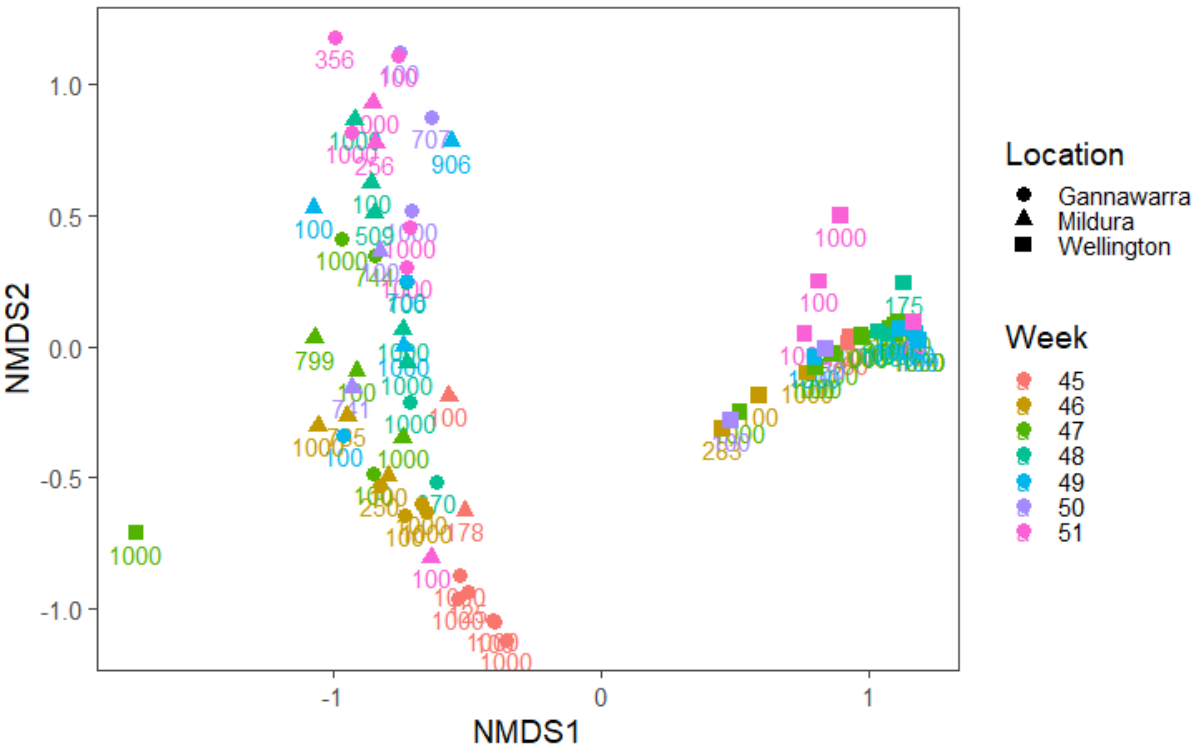


Figure S4: Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity of the (A) viruses and (B) mosquitoes in the trap subsamples from three locations over seven weeks showing a separation between coastal (Wellington) and inland (Gannawarra and Mildura) samples. The number of mosquitoes in each subsample is also shown.

(A)



(B)



CHAPTER 4

Coding-complete genome sequence of Yada Yada virus, a novel alphavirus detected in Australian mosquitoes

4.1 Chapter preface

The metatranscriptomic data produced from field traps in Chapter 3 revealed the presence of a novel alphavirus, which was further characterised in this chapter. Most of the novel viruses detected in Chapter 3 were from insect-specific viral families, which do not pose a risk to public health. However, the *Alphavirus* genus contains viruses pathogenic to humans, warranting investigation into the novel alphavirus detection. The whole genome sequence for the novel alphavirus was assembled from the metatranscriptomic data and used for phylogenetic analysis. The coding-complete genome sequences were made publicly available, and the virus was named Yada Yada virus (YYV).

This chapter is presented in published format, including a correction published after the initial article that details additional information about a virus related to YYV that was not included in the original analysis.

4.2 Publication details

Title: Coding-complete genome sequence of Yada Yada virus, a novel alphavirus detected in Australian mosquitoes

Stage of publication: Published

Journal details: Microbiology Resource Announcements, 2020, volume 9, issue 2, e01476-19. DOI: 10.1128/MRA.01476-19

Authors: Jana Batovska, Jan P. Buchmann, Edward C. Holmes, and Stacey E. Lynch

4.3 Statement of joint authorship

All authors contributed to the conception and design of the experiment; JB performed the nucleic acid extractions, prepared the sequencing libraries, and wrote the manuscript; JB,

ECH, and JPB analysed the data; all authors contributed to the editing of the final manuscript and approved the version submitted for publication.

Statement from co-author confirming the contribution of the PhD candidate:

"As co-author of the manuscript 'Batovska, J., Buchmann, J.P., Holmes, E.C., and Lynch, S.E. (2020). Coding-complete genome sequence of Yada Yada virus, a novel alphavirus detected in Australian mosquitoes. Microbiology Resource Announcements *9*', I confirm that Jana Batovska has made the contributions listed above."

Dr Stacey Lynch

22 June 2020

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Coding-Complete Genome Sequence of Yada Yada Virus, a Novel Alphavirus Detected in Australian Mosquitoes

Jana Batovska,[a,b] Jan P. Buchmann,[c] Edward C. Holmes,[c] Stacey E. Lynch[a]

[a]Agriculture Victoria Research, AgriBio Centre for AgriBioscience, Bundoora, Victoria, Australia
[b]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia
[c]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia

ABSTRACT Here, we report the detection of a novel alphavirus in Australian mosquitoes, provisionally named Yada Yada virus (YYV). Phylogenetic analysis indicated that YYV belongs to the mosquito-specific alphavirus complex. The assembled genome is 11,612 nucleotides in length and encodes two open reading frames.

Alphaviruses (genus *Alphavirus*, family *Togaviridae*) are small (10- to 12-kb) single-stranded positive-sense RNA viruses and include species important to human and animal health, such as *Chikungunya virus* and *Eastern equine encephalitis virus* (1). While these viruses are transmitted primarily by mosquitoes and pathogenic in their vertebrate hosts, there is a small complex of recently discovered alphaviruses that replicate only in mosquito cells (2–5). Here, we report the detection of an alphavirus belonging to this host-restricted complex in the Asia-Pacific region and provide the genome sequence for the novel virus, named Yada Yada virus (YYV).

Virus detection was performed using mosquitoes trapped as part of the Victorian Arbovirus Disease Control Program (6). Encephalitis virus surveillance (EVS) traps (7) were set up overnight each week in three locations in Victoria, Australia, for a total of 7 weeks in late 2016, resulting in 21 trap collections. Traps were sorted into 86 pools of up to 1,000 mosquitoes, which were homogenized in buffer AVL (Qiagen) and centrifuged. RNA was extracted from the supernatant with the QIAamp viral RNA minikit (Qiagen) and used for library preparation, which was performed using the Ovation universal transcriptome sequencing (RNA-Seq) system (NuGEN) with a customized mosquito rRNA depletion (8). The libraries were then treated with free adapter blocking reagent (Illumina) and sequenced on a HiSeq 3000 platform (Illumina) using 2 × 150 bp reads. A total of 909,467,304 paired reads were generated (mean, 10,575,201 per pool; range, 7,971,017 to 16,414,900).

Trinity v2.4.0 (9) was used to trim, normalize, and assemble the reads into contigs, which were taxonomically classified using DIAMOND BLASTx v0.9.22.123 (10) with the NCBI nonredundant (nr) database (acquired 2 September 2019) and an E value cutoff of 10^{-5}. Reads were mapped to assembled contigs using BWA-MEM v0.7.17 r1188 (11). All analyses were performed using default parameters unless stated otherwise. Three of the 21 traps tested contained contigs that had the strongest BLASTx match to the mosquito-specific Eilat alphavirus (EILV). All three traps were collected in November 2016 in Mildura (latitude, 34.249617, longitude, 142.218261). The longest contig was 11,612 nucleotides (nt), with 21-fold average coverage depth and 75.7% amino acid identity to EILV. This contig represents the coding-complete YYV genome, with two open reading frames (ORFs), a 33-nt 3′ leader, a 470-nt 5′ trailer, and 53.4% G+C content. The two ORFs correspond to the structural (1,247 amino acids) and nonstruc-

Citation Batovska J, Buchmann JP, Holmes EC, Lynch SE. 2020. Coding-complete genome sequence of Yada Yada virus, a novel alphavirus detected in Australian mosquitoes. Microbiol Resour Announc 9:e01476-19. https://doi.org/10.1128/MRA.01476-19.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2020 Batovska et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Jana Batovska, jana.batovska@agriculture.vic.gov.au.

Received 28 November 2019
Accepted 13 December 2019
Published 9 January 2020

Downloaded from http://mra.asm.org/ on January 9, 2020 by guest

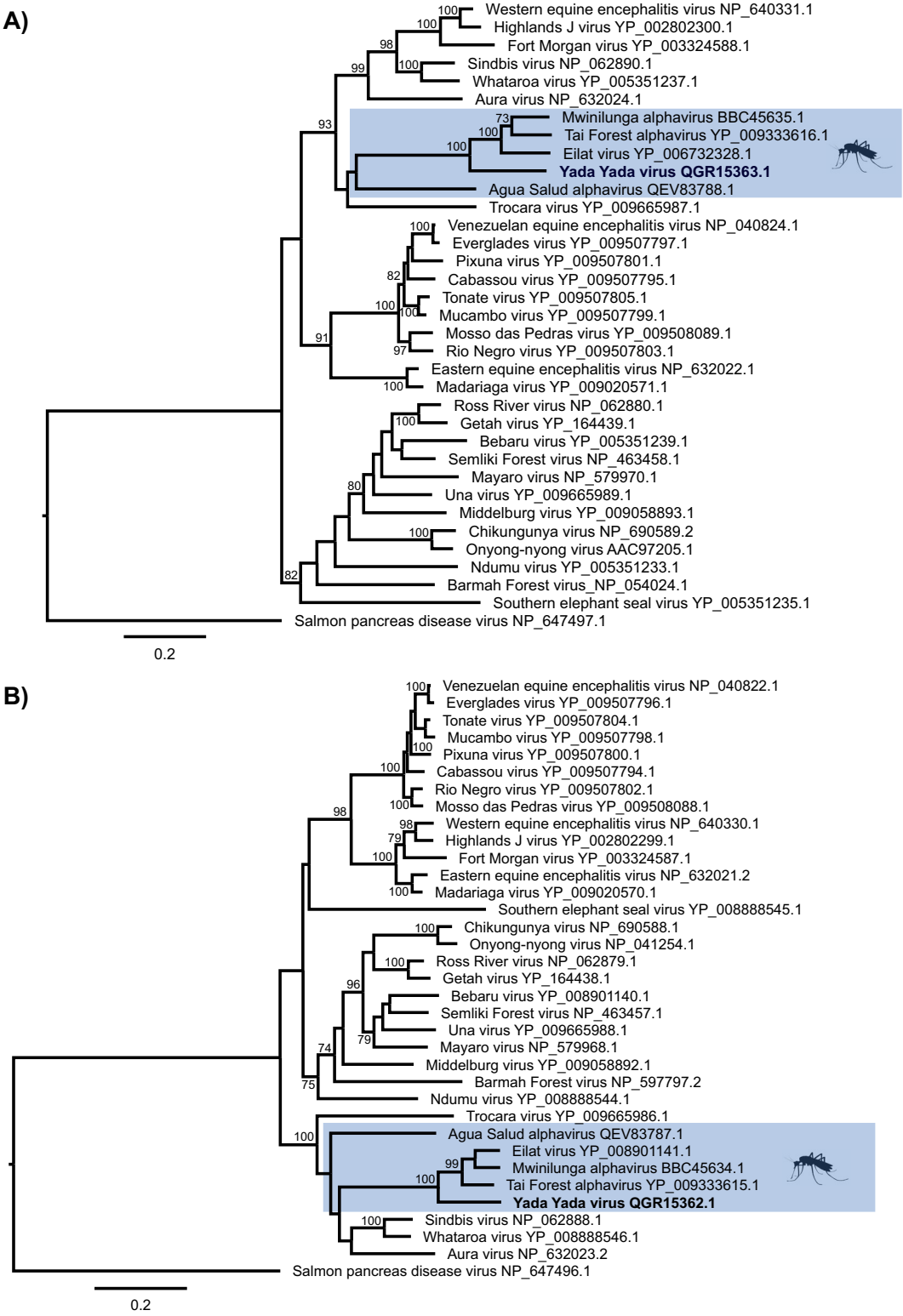


FIG 1 Phylogenetic relationships of YYV and other alphaviruses based on alignments of structural proteins (850 amino acids) (A) and nonstructural proteins (1,287 amino acids) (B). Maximum likelihood trees were estimated using the LG plus gamma model of amino acid substitution in PhyML, with 1,000 bootstrap replicates, and rooted using salmon pancreas disease virus. Bootstrap values greater than 70% are shown beside the branches, and the GenBank accession numbers are shown with the virus names. The mosquito-specific complex is highlighted in blue with YYV in bold.

Downloaded from http://mra.asm.org/ on January 9, 2020 by guest

tural (2,437 amino acids) proteins. Translation of the genome sequence was performed using the ExPASy Translate tool (12).

Phylogenetic analysis was performed by the creation of alignments of YYV and the structural and nonstructural protein sequences of other alphaviruses using MAFFT v7.429 (13), the removal of ambiguously aligned residues with TrimAl v1.4.1 (14), and maximum likelihood inference using PhyML v3.1 (15) employing the Le-Gascuel (LG) plus gamma distribution model of amino acid substitution and 1,000 bootstrap replicates. The resultant trees were then viewed in FigTree v1.4.4 (16). In both the structural (Fig. 1A) and nonstructural (Fig. 1B) protein trees, YYV was placed in the mosquito-specific alphavirus complex, suggesting that it might also have a restricted host range.

To investigate the vector of YYV, the assembled contigs were compared to a cytochrome oxidase I (COI) database of Australian mosquito species (17) using BLASTn v2.9.0+ (18) with an E value cutoff of 10^{-5}, and the results were filtered for alignments >200 bp in length and matches of >95% identity. Only two mosquito species were present in all three traps, *Anopheles annulipes* and *Culex australicus*/*Culex globocoxitus* (these two *Culex* species are indistinguishable using COI), supporting previous studies that have detected mosquito-specific alphaviruses from only *Anopheles* and *Culex* species (2–5). Due to the homogenization of the traps, the YYV vector species cannot be definitively determined. However, read mapping showed that the abundance of *A. annulipes* was associated with YYV genome coverage, whereas the abundance of *C. australicus*/*C. globocoxitus* was not (data not shown).

The discovery of YYV expands the diversity and geographic range of the mosquito-specific alphavirus complex and in doing so will help reveal the virus origin and evolution of host switching (19). In addition, it is noteworthy that mosquito-specific viruses that are closely related to pathogenic vertebrate viruses have potential applications in vaccine development and as biocontrol agents (20).

Data availability. The YYV genome sequence has been deposited in GenBank under the accession number MN733821. The sequencing reads are available in the SRA database via BioProject accession number PRJNA594295.

ACKNOWLEDGMENTS

We thank Dale Hutchinson for conducting the trapping and Karen Brown for her assistance in processing the mosquitoes.

This work was funded by the Biosciences Research Innovation Fund Program provided by the Victorian Department of Jobs, Precincts, and Regions. The Victorian Arbovirus Disease Control Program is funded by the Department of Health and Human Services. J.B. is supported by an Australian Government research training program scholarship. E.C.H. is supported by an Australian Research Council Australian Laureate fellowship (FL170100022).

REFERENCES

1. Chen R, Mukhopadhyay S, Merits A, Bolling B, Nasar F, Coffey LL, Powers A, Weaver SC, ICTV Report Consortium. 2018. ICTV virus taxonomy profile: *Togaviridae*. J Gen Virol 99:761–762. https://doi.org/10.1099/jgv.0.001072.
2. Hermanns K, Zirkel F, Kopp A, Marklewitz M, Rwego IB, Estrada A, Gillespie TR, Drosten C, Junglen S. 2017. Discovery of a novel alphavirus related to Eilat virus. J Gen Virol 98:43–49. https://doi.org/10.1099/jgv.0.000694.
3. Nasar F, Palacios G, Gorchakov RV, Guzman H, Da Rosa APT, Savji N, Popov VL, Sherman MB, Lipkin WI, Tesh RB, Weaver SC. 2012. Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. Proc Natl Acad Sci U S A 109:14622–14627. https://doi.org/10.1073/pnas.1204787109.
4. Hermanns K, Marklewitz M, Zirkel F, Overheul GJ, Page RA, Loaiza JR, Drosten C, van Rij RP, Junglen S. 2019. Agua Salud alphavirus defines a novel lineage of insect-specific alphaviruses discovered in the New World. J Gen Virol. https://doi.org/10.1099/jgv.0.001344.
5. Torii S, Orba Y, Hang'ombe BM, Mweene AS, Wada Y, Anindita PD, Phongphaew W, Qiu Y, Kajihara M, Mori-Kajihara A, Eto Y, Harima H, Sasaki M, Carr M, Hall WW, Eshita Y, Abe T, Sawa H. 2018. Discovery of Mwinilunga alphavirus: a novel alphavirus in *Culex* mosquitoes in Zambia. Virus Res 250:31–36. https://doi.org/10.1016/j.virusres.2018.04.005.
6. Wong S, Brown K, Crowder J, Chea S, Mee P, Batovska J, Lynch S. 2017. Victorian Arbovirus Disease Control Program annual report 2016–2017. Agriculture Victoria, Victoria, Australia.
7. Ritchie SA, Kline DL. 1995. Comparison of CDC and EVS light traps baited with carbon dioxide and octenol for trapping mosquitoes in Brisbane, Queensland (Diptera: Culicidae). Aust J Entomol 34:215–218. https://doi.org/10.1111/j.1440-6055.1995.tb01322.x.
8. Batovska J, Mee PT, Lynch SE, Sawbridge TI, Rodoni BC. 2019. Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool. Sci Rep 9:19398. https://doi.org/10.1038/s41598-019-55741-3.
9. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from

Downloaded from http://mra.asm.org/ on January 9, 2020 by guest

RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883.
10. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
11. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN]. https://arxiv.org/abs/1303.3997.
12. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788. https://doi.org/10.1093/nar/gkg563.
13. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
14. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
15. 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. https://doi.org/10.1093/sysbio/syq010.
16. Rambaut A. 2019. FigTree v1.4: tree figure drawing tool. GitHub. https://github.com/rambaut/figtree/releases.
17. Batovska J, Blacket MJ, Brown K, Lynch SE. 2016. Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. Ecol Evol 6:3001–3011. https://doi.org/10.1002/ece3.2095.
18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
19. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, Qin X-C, Xu J, Holmes EC, Zhang Y-Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife 4. https://doi.org/10.7554/eLife.05378.
20. Hall RA, Hobson-Peters J. 2018. Newly discovered mosquito viruses help control vector-borne viral diseases. Microbiol Aust 39:72–75. https://doi.org/10.1071/MA18020.

Downloaded from http://mra.asm.org/ on January 9, 2020 by guest

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

AUTHOR CORRECTION

Correction for Batovska et al., "Coding-Complete Genome Sequence of Yada Yada Virus, a Novel Alphavirus Detected in Australian Mosquitoes"

Jana Batovska,[a,b] Jan P. Buchmann,[c] Edward C. Holmes,[c] Stacey E. Lynch[a]

[a]Agriculture Victoria Research, AgriBio Centre for AgriBioscience, Bundoora, Victoria, Australia

[b]School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, Australia

[c]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia

Volume 9, no. 2, e01476-19, 2020, https://doi.org/10.1128/MRA.01476-19. After publication, the authors were made aware of a short sequence on GenBank (JQ749729.1) with high similarity to Yada Yada virus (YYV). The 331-bp sequence is part of the nonstructural polyprotein (NSP4) and shares 97% nucleotide identity and 99.1% amino acid identity with YYV. This sequence was detected in an *Aedes notoscriptus* mosquito collected in North Melbourne, Australia, in 2011. The similarity of this sequence to YYV suggests that the virus from which it originated is part of the mosquito-specific alphavirus complex, which consists of viruses that have been detected only in *Anopheles* and *Culex* mosquitoes.

Citation Batovska J, Buchmann JP, Holmes EC, Lynch SE. 2020. Correction for Batovska et al., "Coding-complete genome sequence of Yada Yada virus, a novel alphavirus detected in Australian mosquitoes." Microbiol Resour Announc 9:e00103-20. https://doi.org/10.1128/MRA.00103-20.

Copyright © 2020 Batovska et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Jana Batovska, jana.batovska@agriculture.vic.gov.au.

Published 5 March 2020

Downloaded from http://mra.asm.org/ on March 8, 2020 by guest

CHAPTER 5

Metagenomic arbovirus detection using MinION nanopore sequencing

5.1 Chapter preface

To further explore how metatranscriptomics can be applied arbovirus surveillance, this chapter details the first time MinION nanopore sequencing was used to metatranscriptomically detect an arbovirus from a mosquito sample. The performance of the MinION was compared to the commonly used MiSeq sequencer. The information produced in this chapter provides insights into the utility of the MinION for arbovirus surveillance, with its portability, low cost, and real-time long read sequencing offering enhanced, in-field surveillance capabilities.

This chapter is presented in published format. Chronologically, this chapter is the first of the thesis, and uses the term 'metagenomic' instead of 'metatranscriptomic'. While the two are used interchangeably in the literature, 'metatranscriptomic' is specific to RNA sequencing, and so was later switched to better describe the methods used in this thesis.

5.2 Publication details

Title: Metagenomic arbovirus detection using MinION nanopore sequencing

Stage of publication: Published

Journal details: Journal of Virological Methods, 2017, volume 249, pages 79–84. DOI: 10.1016/j.jviromet.2017.08.019

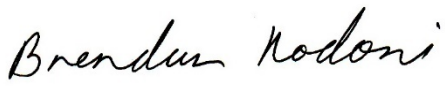
Authors: Jana Batovska, Stacey E. Lynch, Brendan C. Rodoni, Tim I. Sawbridge, and Noel O.I. Cogan

5.3 Statement of joint authorship

All authors contributed to the conception and design of the experiment; JB performed the nucleic acid extractions, prepared the sequencing libraries, analysed the data, and wrote the manuscript; all authors contributed to the editing of the final manuscript and approved the version submitted for publication.

Statement from co-author confirming the contribution of the PhD candidate:

“As co-author of the manuscript ‘Batovska, J., Lynch, S.E., Rodoni, B.C., Sawbridge, T.I., and Cogan, N.O. (2017). Metagenomic arbovirus detection using MinION nanopore sequencing. Journal of Virological Methods *249*, 79–84’, I confirm that Jana Batovska has made the contributions listed above.”



Professor Brendan Rodoni

22 June 2020

Journal of Virological Methods 249 (2017) 79–84

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Metagenomic arbovirus detection using MinION nanopore sequencing

Jana Batovska[a,b,*], Stacey E Lynch[a], Brendan C Rodoni[a,b], Tim I Sawbridge[a,b], Noel OI Cogan[a,b]

[a] Agriculture Victoria Research, AgriBio Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3083, Australia
[b] School of Applied Systems Biology, La Trobe University, Bundoora, Victoria, 3086, Australia

ARTICLE INFO

Keywords:
Mosquito
Culicidae
Ross River virus
MinION
MiSeq
Surveillance

ABSTRACT

With its small size and low cost, the hand-held MinION sequencer is a powerful tool for in-field surveillance. Using a metagenomic approach, it allows non-targeted detection of viruses in a sample within a few hours. This study aimed to determine the ability of the MinION to metagenomically detect and characterise a virus from an infected mosquito. RNA was extracted from an *Aedes notoscriptus* mosquito infected with Ross River virus (RRV), converted into cDNA and sequenced on the MinION. Bioinformatic analysis of the MinION reads led to detection of full-length RRV, with reads of up to 2.5 kb contributing to the assembly. The cDNA was also sequenced on the MiSeq sequencer, and both platforms recovered the RRV genome with > 98% accuracy. This proof of concept study demonstrates the metagenomic detection of an arbovirus, using the MinION, directly from a mosquito with minimal sample purification.

1. Introduction

Metagenomic next-generation sequencing (NGS) allows for the unbiased detection of organisms within a sample. This powerful approach has enhanced pathogen detection for both diagnostic and surveillance applications (Barzon et al., 2011; Temmam et al., 2014). Until recently, NGS was restricted to the laboratory due to the size of the sequencers available; however, the release of the MinION (Oxford Nanopore Technologies) enables NGS in-field. The MinION is a low-cost, hand-held sequencer that produces long reads (up to 233 kb) in real-time (Jansen et al., 2017). At time of writing, the cost of a starter kit that includes the sequencer is USD $1000, with additional kits and flow cells ranging from USD $500 to $900. The sequencing accuracy is considerably lower than popular short-read sequencers such as the Illumina MiSeq and HiSeq, with an error rate of approximately 5–10% (Tyson et al., 2017) for 2D R9 reads, compared to < 0.1% for the Illumina sequencers (Houldcroft et al., 2017). However, multi-fold genome coverage can be used to generate an accurate consensus sequence, with the MinION achieving > 99% accuracy post-data analysis (Wang et al., 2015). Furthermore, low quality reads can often be adequate to identify a known pathogen (Walter et al., 2016), allowing rapid detection with more accurate characterisation performed later.

The MinION has been used to detect a variety of viruses, including Ebola (Hoenen et al., 2016; Quick et al., 2016), dengue (Mongan et al., 2015), Zika (Quick et al., 2017), influenza (Eckert et al., 2016; Wang et al., 2015), Flock House (Jaworski and Routh, 2017), and cowpox (Kilianski et al., 2015). The MinION has a lower throughput than NGS platforms such as Illumina (Lu et al., 2016), so these studies targeted the virus using PCR amplicons or hybridisation capture. This approach limits detection to known pathogens with characterised genome sequences. Despite a lower throughput, the MinION has been used to sequence Ebola, chikungunya, and hepatitis C viruses using an unbiased metagenomic approach with randomly amplified cDNA (Greninger et al., 2015). Kilianski et al. (2016) also used a metagenomic approach when sequencing Ebola virus and Venezuelan equine encephalitis virus with the MinION, however unamplified RNA/cDNA-hybrids were used for library preparation. Both of these studies indicate the MinION is sensitive enough to detect viruses without target enrichment. While all of the viruses sequenced using nanopore technology have been from clinical samples or viral cultures, the MinION has yet to be used to detect viruses directly from a viral vector, such as a mosquito.

Aedes albopictus and *Ae. aegypti* are key vectors of medically important arboviruses such as dengue, chikungunya and Zika (Patterson et al., 2016). As invasive species, they readily colonise new territories (Medlock et al., 2012; Zhong et al., 2013) and have been implicated in importing arboviruses into non-endemic areas (Lindsay et al., 2015). In efforts to prevent their establishment, biosecurity surveillance programs have included exotic mosquito monitoring at high-risk international ports for early detection (Knope et al., 2016; Medlock et al., 2012; Vaux and Medlock, 2015). Exotic mosquito surveillance relies on morphological identification of collected specimens, and is often complemented by molecular approaches, such as PCR, to confirm mosquito

* Corresponding author at: Agriculture Victoria Research, AgriBio Centre for AgriBioscience, 5 Ring Road, Bundoora, Victoria, 3083, Australia.
E-mail addresses: jana.batovska@ecodev.vic.gov.au (J. Batovska), stacey.lynch@ecodev.vic.gov.au (S.E. Lynch), brendan.rodoni@ecodev.vic.gov.au (B.C. Rodoni), tim.sawbridge@ecodev.vic.gov.au (T.I. Sawbridge), noel.cogan@ecodev.vic.gov.au (N.O. Cogan).

http://dx.doi.org/10.1016/j.jviromet.2017.08.019
Received 12 June 2017; Received in revised form 22 August 2017; Accepted 23 August 2017
0166-0934/ © 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

species or screen for specific arboviruses. These approaches can take several days and delay appropriate control measures. The MinION can bring molecular testing of invasive species to the port of entry, and identify vector species and arbovirus carriage in real-time.

The aim of this study was to determine if MinION nanopore sequencing could be used to detect an arbovirus from a single mosquito using a metagenomic approach. In order to provide the most utility in biosecurity situations, the ability to detect the species of the mosquito was also tested. The performance of the MinION in this study was directly compared to the commonly used, lab-based MiSeq.

2. Materials and methods

2.1. Nucleic acid extraction and cDNA synthesis

A single *Ae. notoscriptus* mosquito infected with Ross River virus (RRV) strain QML1 (Jones et al., 2010) was used as the test sample. Rearing and infection of the mosquito is described in Batovska et al. (2017). The mosquito was homogenised in 180 µL of viral lysis buffer (Buffer AVL, Qiagen) using a TissueLyser (Qiagen) at 30 Hz and 2 cycles of 1 min. The homogenate was centrifuged (5 min, 10,000 rpm), and 140 µL of the supernatant was used for viral RNA purification using the QIAamp Viral RNA Mini Kit (Qiagen). The kit was used according to the manufacturer's instructions, however no carrier RNA was added to the Buffer AVL. The purified RNA was eluted in 80 µL and DNase treated using a TURBO DNA-free Kit (Ambion) according to manufacturer's instructions.

The RNA extract was quantified by a 2200 TapeStation using the RNA ScreenTape assay (Agilent Technologies). The quantification resulted in 10 µL of RNA being used as input for first-strand cDNA synthesis. The RNA was reverse transcribed with the SuperScript III First-Strand Synthesis System (Invitrogen) using 100 ng of random hexamers. Second strand cDNA synthesis was performed with the NEBNext mRNA Second Strand Synthesis Module (New England BioLabs). The cDNA was purified using AMPure XP beads (Beckman Coulter) with a 1.8 x beads ratio, quantified by a Qubit 1.0 Fluorometer (Life Technologies), and stored at −20 °C.

2.2. Quantification of viral load

The viral load of the mosquito was quantified using droplet digital PCR (ddPCR). The primer pair RRVE2F and RRVE2R, and RRVE2 Probe were used to amplify and detect RRV (Hall et al., 2011). The reaction mixture consisted of 12 µL of 2 x ddPCR Supermix for Probes (No dUTP) (Bio-Rad), 0.12 µL of each 100 µM/L primer, 0.06 µL of 100 µM/L probe, 0.03 µL of 100,000 U/mL HindIII-HF (New England Biolabs), 10.7 µL of water, and 6.25–200 pg of cDNA. The reaction mixture was partitioned into droplets using the Automated Droplet Generator loaded with oil for probes (Bio-Rad). After droplet generation, PCR was performed using the following conditions: 95 °C for 10 min; 40 cycles of 94 °C for 30 s and 60 °C for 1 min; and 98 °C for 10 min. A 2 °C/sec ramp rate was used for each step. The droplets were then analysed with a QX200 Droplet Reader (Bio-Rad) and QuantaSoft™ Analysis Pro software was used to provide the number of RRV cDNA copies per µL of the ddPCR reaction. A positive and negative control was used in the quantification reaction to verify results.

2.3. Nanopore library preparation and sequencing

Libraries were prepared with the SQK-NSK007 Sequencing Kit (R9 version) using the Oxford Nanopore Technologies (ONT) 2D cDNA sequencing protocol. The undiluted cDNA was A-tailed using the NEBNext Ultra II End Repair/dA-Tailing Module (New England BioLabs). The incubation periods were extended to 30 min to improve A-tailing. Adapters (ONT) were ligated onto the cDNA, and PCR was performed using LongAmp *Taq* DNA Polymerase (New England BioLabs), primers (ONT) and a 10 min 65 °C extension step. The amplified cDNA was again A-tailed using the method specified above, and ONT hairpin adapters and tethers were ligated onto the cDNA. Library purification was performed using Dynabeads MyOne Streptavidin C1 beads (Invitrogen) with a 0.5 x beads ratio, and a 30 min incubation at room temperature. The eluted library was quantified by a Qubit, and sequenced on a primed MinION Spot-ON Flow Cell Mk I (R9 version). The NC_48Hr_Sequencing_Run_FLO_MAP103 program was run on MinKNOW with local basecalling, thereby producing 1D reads. The use of 1D reads was chosen to determine if the field-appropriate Rapid 1D Sequencing Kit would be adequate for virus detection.

2.4. Illumina library preparation and sequencing

The cDNA was diluted to 0.2 ng/µl and a total of 1 ng was used for library preparation, performed as per the Nextera XT DNA Sample Preparation Kit protocol (Illumina). The resulting library was quantified and evaluated for fragment size using a 2200 TapeStation (Agilent Technologies), diluted to 10 pM and sequenced on an Illumina MiSeq platform (2 × 250 bp reads).

2.5. Data analysis

The MinION HDF5 files were processed using poretools version 0.6.0 (Loman and Quinlan, 2014) to extract read data in FASTA format. Adapters were removed from the MinION reads using Cutadapt version 1.9 (Martin, 2011). The trimmed MinION reads were aligned to the RRV strain QML1 genome reference sequence (GQ433354.1) using BWA-MEM version 0.7.7 (Li, 2013) with default parameters. The polyA tail was removed from the RRV reference sequence to improve reliability of the alignment. Coverage was calculated using SAMtools version 0.1.19 (Li et al., 2009), and similarity was calculated by determining the number of matches/mismatches at each base using pysamstats version 0.24.3 (available at: https://github.com/alimanfoo/pysamstats, accessed 10 January 2016). Variant calling was performed with BCFtools in the SAMtools package, with a quality threshold set at 200. The consensus sequence was derived and aligned to the RRV reference sequence using ClustalW in Geneious version 8.1.8 (Kearse et al., 2012). Taxonomic classification of the trimmed MinION reads was performed using a BLASTn version 2.3.0 search with the NCBI nucleotide database and the following parameters: e-value 1×10^{-5}; word size 28; and a maximum of one hit per read. Mosquito species identification was also performed using BLASTn with the same parameters and a custom database of *Cytochrome oxidase I* (COI) sequences from Australian mosquito species (Batovska et al., 2016).

The demultiplexed MiSeq reads were trimmed of adapters and bad quality bases using Trim Galore version 0.4.0 (Krueger, 2015). Reads were trimmed of bases with a PHRED quality score < 20 and were removed if they were < 100 bp long. In order to have comparable data, the trimmed MiSeq reads were normalised so that the total number of base pairs generated was similar to the total number of MinION base pairs. The normalised MiSeq reads were then analysed using the same method as for the MinION data.

The unprocessed MinION HDF5 files (SRR5572188) and MiSeq FASTQ files (SRR5572189) used in this study have been deposited into the National Centre for Biotechnology Information (NCBI) Sequence Read Archive under project ID PRJNA386415.

3. Results

3.1. Virus sequencing using the MinION

The viral load of the RRV-infected mosquito was estimated at 1.17×10^4 cDNA copies/µL using ddPCR.

When sequenced on the MinION, a total of 229 reads were identified as originating from RRV, accounting for 0.28% of the total reads

Table 1
Sequencing run data and Ross River virus (RRV) reference alignment statistics for the MiSeq and MinION. The mean read length is post-adapter trimming. The MiSeq raw reads are unpaired have been normalised according to total base pairs (bp).

	Raw reads	Total bp	Mean read length (bp)	Mapped reads to RRV (%)	Mean RRV coverage (x)	Pairwise identity (%)	% of ref. seq.
MinION	82,259	89,244,295	1,084.9	0.28	10	98.1	98.9
MiSeq	467,756	89,244,548	190.8	0.41	62	98.3	97.8

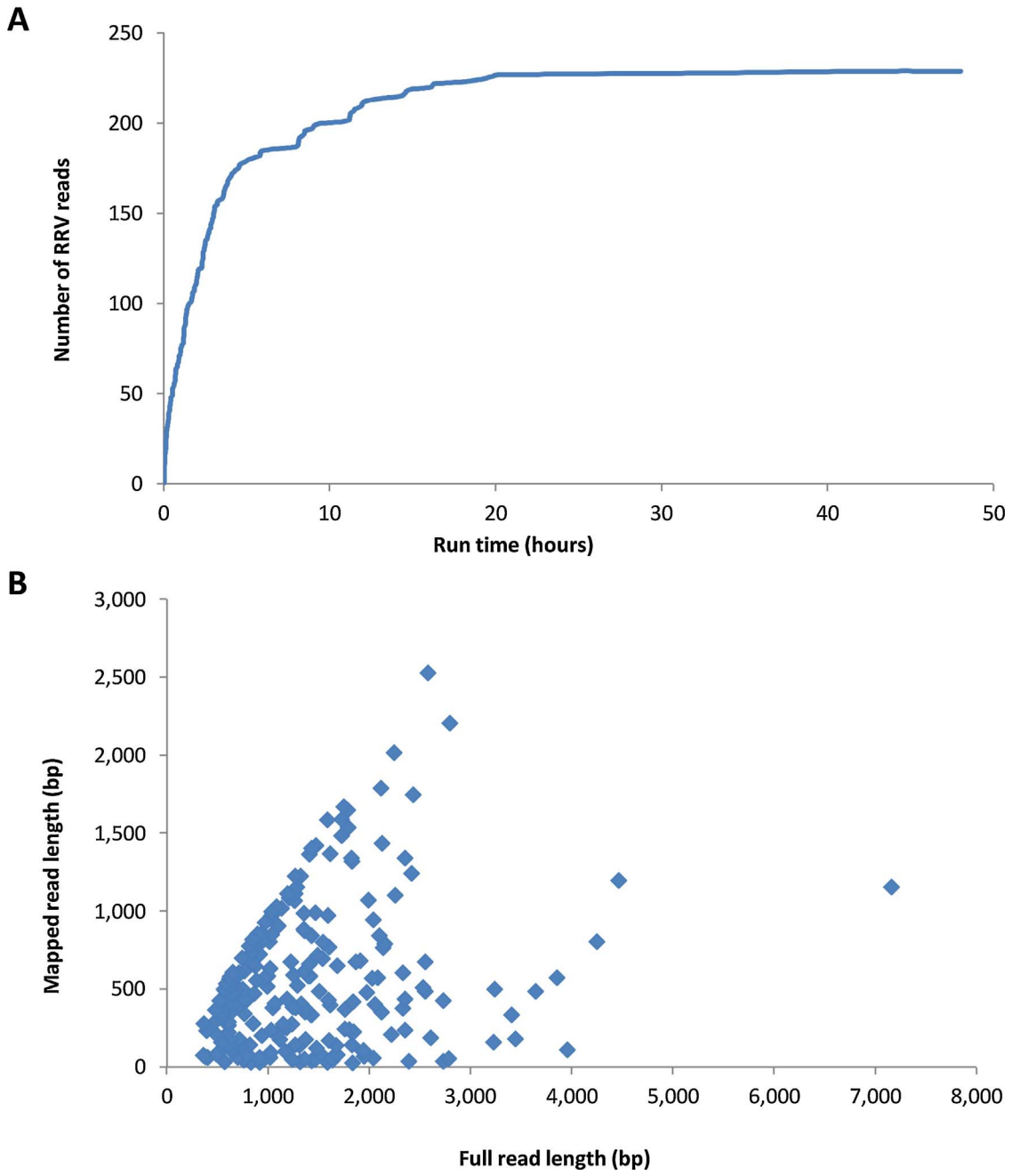


Fig. 1. Descriptive statistics of Ross River virus (RRV) reads generated by the MinION. A) The cumulative number of RRV reads generated throughout the 48 h sequencing run. B) A comparison of the length of the MinION reads that were mapped to the Ross River virus (RRV) reference sequence, compared to the original full length of the reads.

(Table 1). After 10 min of sequencing 12% of the RRV reads had been generated, increasing to 32.3% at one hour, and after 10 h the majority of RRV reads (87.3%) were sequenced (Fig. 1A). The generation of the RRV reads was proportional to the total number of reads generated during the course of the MinION run.

The MinION produced a range of RRV read lengths (mean: 1388.4 bp; range: 355–7159 bp); however, often large regions of the RRV reads produced by the MinION did not align to the reference sequence (Fig. 1B). A mean of 45.6% of the MinION read length aligned to the reference, with the remainder being clipped by the BWA-MEM aligner. Upon manual inspection, it was found that the clipped regions had higher levels of sequencing error compared to the mapped regions. After clipping, the longest RRV read length was 2528 bp and the mean RRV read length was 547.8 bp.

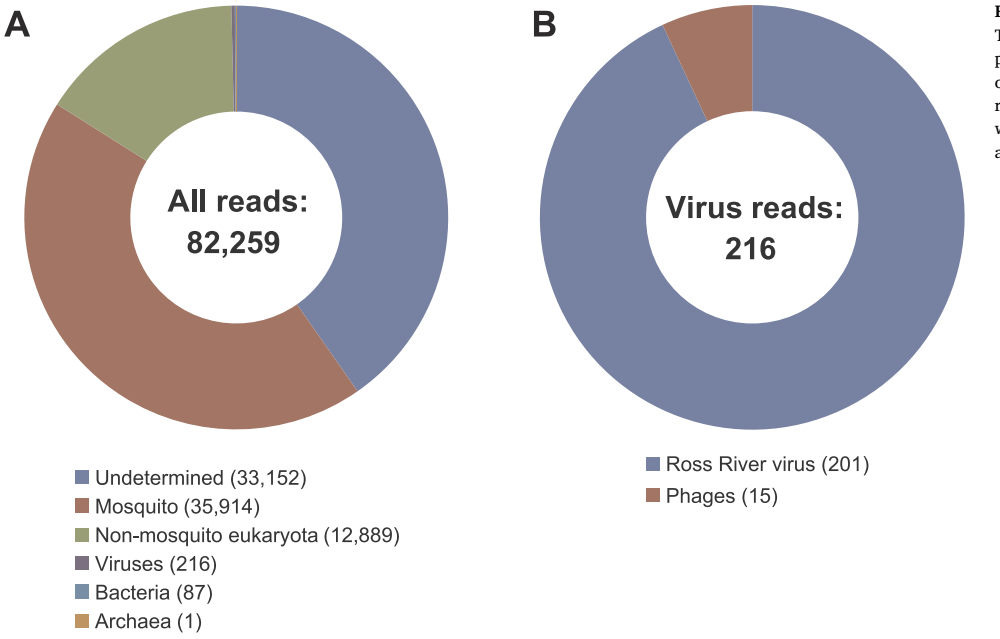


Fig. 2. BLASTn analysis of the MinION reads. A) Taxonomic classification of all the MinION reads, post adapter trimming. B) Taxonomic classification of the reads identified as virus. The total number of reads used is shown in the centre of the donut, whereas the number of reads assigned to each taxon are shown in brackets below the donut.

3.2. Taxonomic classification of the MinION reads

A BLASTn search using the NCBI nucleotide database allowed for characterisation of the MinION reads and untargeted detection of RRV. The majority of reads (59.3%) were classified as eukaryota, with most of these belonging to mosquito (Fig. 2A). Of the non-mosquito eukaryota reads, 79.9% were assigned to nematodes. The finding of nematode sequences was supported by the Illumina data, with 95.5% of non-mosquito eukaryote MiSeq reads belonging to nematodes. The majority of these reads belonged to a 4.2 kB region of the *Brugia timori* assembly (GenBank accession LK907835.1; 87.6% and 72.1% of MinION and MiSeq reads respectively). Of the 216 MinION reads that were identified as virus, 93.1% were RRV, with the rest classified as phage-related (Fig. 2B).

In addition to virus detection, the species of the host mosquito was successfully derived from the MinION data, with 100% of the mosquito COI BLASTn results matching *Ae. notoscriptus* (total = 24 reads).

3.3. MinION vs. MiSeq

The MinION generated poorer quality RRV reads than the MiSeq, with a 4.7% error rate compared to 0.7% for the MiSeq (Fig. 3A). The MinION also generated less RRV reads than the MiSeq (229 vs. 1919 reads), giving it less coverage of the genome (10-fold vs. 62-fold; Fig. 3B). However, the MinION RRV reads were longer than those produced by the MiSeq, with a mean RRV read length of 547.8 bp (range: 30–2528 bp) compared to 187.6 bp (range: 30–250 bp).

When comparing consensus RRV genome sequences, the performance of the MinION and MiSeq were comparable, with both platforms acquiring a similar percentage of the reference genome (98.9 vs. 97.8%), and similar pairwise identity (98.1 vs. 98.3%; Table 1).

Variant calling revealed a homozygous single nucleotide polymorphism (SNP) from the RRV reference genome at 10,065 bp. The SNP was present in both the MinION and MiSeq data (Fig. 3A).

4. Discussion

This study demonstrates the ability of the MinION to detect and sequence an arbovirus directly from an infected mosquito. The full-length RRV genome was recovered from the MinION reads, and a BLASTn search revealed the presence of RRV without targeting known viruses (Fig. 2). The BLASTn search also detected the presence of other organisms such as bacteria and nematodes, highlighting the MinION's potential to detect all the pathogens present within a mosquito. Due to the limited coverage of the nematode genome, the source of the nematode sequences requires further investigation. It should be noted that pathogen detection using metagenomic approaches is limited by the sequence databases available for classification.

The MinION is able to perform pathogen detection rapidly, with 12% of the RRV reads produced within the first 10 min of the run (Fig. 1A). Due to the MinION's real-time sequencing, reads can be analysed as they are generated, significantly speeding up virus detection (Walter et al., 2016). The ability to rapidly detect pathogens directly from a mosquito combined with the MinION's portability and affordability make it ideal for in-field biosecurity surveillance.

The utility of the MinION in biosecurity is further demonstrated by its ability to determine the species of the mosquito, in addition to detecting viruses. Despite being DNase-treated, reads specific to *Ae. notoscriptus* were present in the MinION data, leading to the correct identification of the mosquito host. Invasive mosquitoes such as *Ae. albopictus* are intercepted at international ports throughout the year, making species identification a vital aspect of biosecurity surveillance (Vaux and Medlock, 2015). Current methods involve morphological identification, which can be unreliable and difficult to perform on immature specimens. Molecular identification of species using a portable sequencer would allow decisions about vector control to be made quickly and help prevent the establishment of exotic mosquitoes.

This study supports previous findings (Greninger et al., 2015; Wang et al., 2015) that indicate virus characterisation using the MinION is comparable to that of the MiSeq. While the MinION had a higher error rate and lower coverage (Fig. 3), the pairwise identity and length of the two RRV consensus genomes was similar (98.1 vs. 98.3%; Table 1). Furthermore, even with a high error rate, the raw MinION reads could still be used for accurate virus detection (Fig. 2). The longer reads produced by nanopore sequencing facilitate the ability to accurately identify viruses, with reads as long as 2.5 kb mapping to the RRV genome. However, many of the mapped RRV MinION reads had regions clipped by the BWA-MEM aligner due to high error rate (Fig. 1B). Clipping could be reduced by error correcting the reads first with programs such as Nanocorrect (Loman et al., 2015). Alternatively, an aligner that does not clip reads could be used to analyse the data, although this could lead to decreased pairwise identity. Nevertheless,

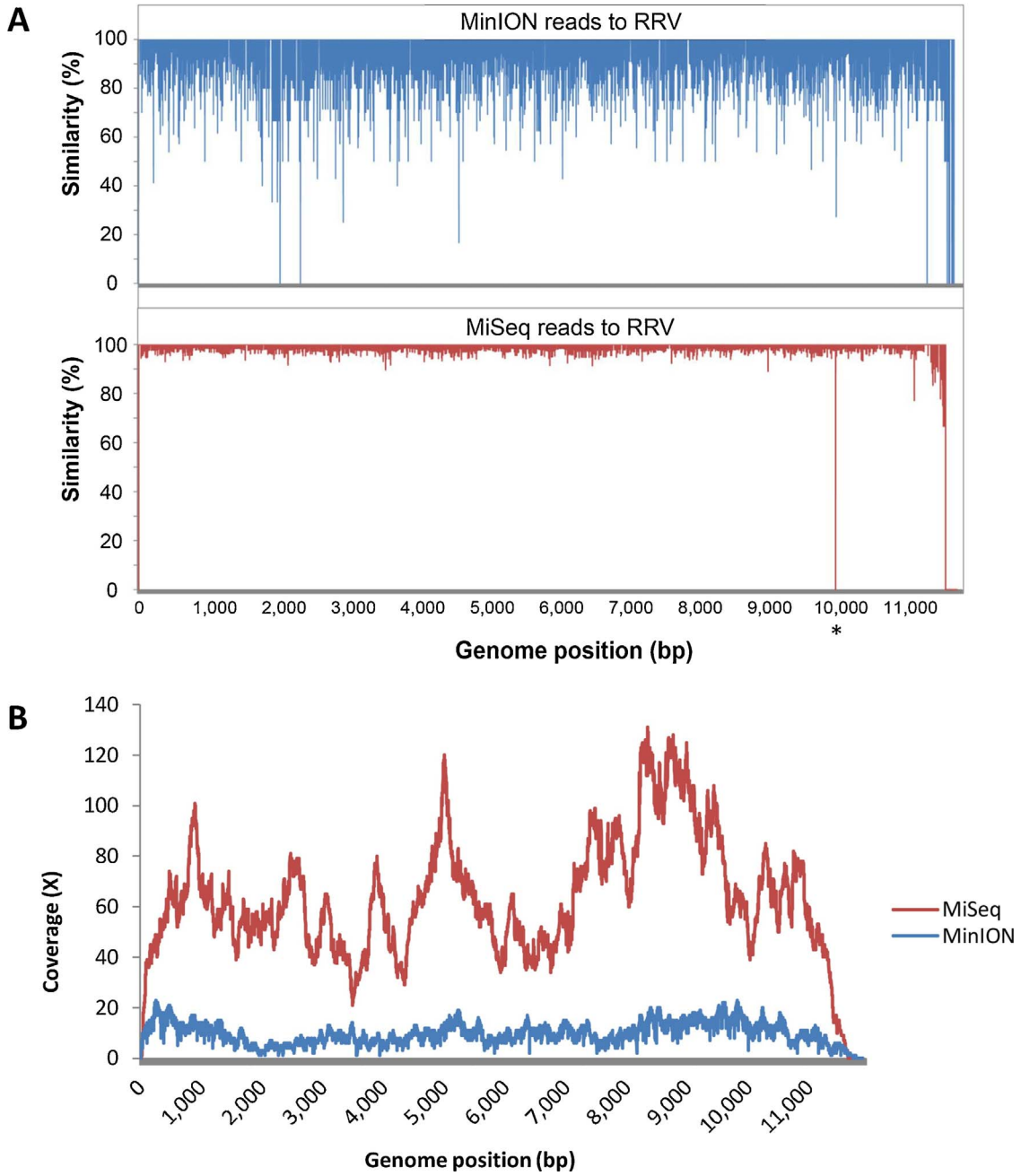


Fig. 3. Comparison of the A) similarity and B) coverage of MinION and MiSeq reads when aligned to the Ross River virus (RRV) reference sequence. The * symbol in **A** indicates the position of a single nucleotide polymorphism (SNP).

using BWA-MEM with uncorrected MinION reads remains sufficient for virus identification and directly comparable with short-read sequencing technology.

Now that the capacity of the MinION to sequence virus from a mosquito sample has been established, future studies can test the method in-field. A range of protocols and equipment have been designed to mobilise laboratory procedures and shorten turnaround time. For instance, the VolTRAX library preparation system, and the prospective SmidgION (Lu et al., 2016). However, the quality of the data produced by these devices has yet to be established. Improvements in bioinformatic processing also assist with in-field use, with the recent release of local basecallers, such as Nanocall (David et al., 2017), removing the need for an internet connection. Currently, the primary restriction to local basecalling is the large amount of computing power required; however, this will likely improve with time (Walter et al., 2016).

In-field use of the MinION would be further enabled with the Rapid 1D Sequencing Kit, which reduces library preparation time to approximately 10 min (Mitsuhashi et al., 2017). The results of this study indicate 1D reads are of sufficient quality for virus detection in mosquitoes. The direct sequencing of RNA would also reduce sample processing time, and has recently been demonstrated with the MinION (Garalde et al., 2016). The method currently applies only to polyadenylated RNA, but future advancements could expand this capability to all RNA and help avoid the biases associated with cDNA synthesis.

This study serves as a proof of concept for the metagenomic detection of an arbovirus from a mosquito sample using a MinION sequencer. This method of pathogen detection and characterisation differs from previous nanopore sequencing studies in that it uses the mosquito as the sample rather than clinical samples or viral cultures,

thereby providing a rapid and unbiased surveillance tool. Furthermore, no target enrichment is used, allowing virus detection without a priori knowledge. The MinION is affordable, portable, and able to detect arbovirus quickly and with sufficient accuracy, demonstrating its potential for in-field virus surveillance. Utilisation of recent protocols and device releases could allow virus detection on the MinION to be performed within a few hours in-field. Nanopore sequencing technology is advancing at an unprecedented pace, promising a future where portable sequencing will be routine in surveillance and many other fields.

Acknowledgements

The authors thank Jonathan Darbro and Elise Kho for providing the mosquito specimen used in this study, and Simone Warner for critically reading and approving the manuscript. This project was funded through a Biosciences Research Innovation Fund Program provided by the Department of Economic Development, Jobs, Transport and Resources.

References

Barzon, L., Lavezzo, E., Militello, V., Toppo, S., Palù, G., 2011. Applications of next-Generation sequencing technologies to diagnostic virology. Int. J. Mol. Sci. 12, 7861–7884.

Batovska, J., Blacket, M.J., Brown, K., Lynch, S.E., 2016. Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. Ecol. Evol. 6, 3001–3011.

Batovska, J., Lynch, S.E., Cogan, N.O.I., Brown, K., Darbro, J.M., Kho, E.A., Blacket, M.J., 2017. Effective mosquito and arbovirus surveillance using metabarcoding. Mol. Ecol. Resour. http://dx.doi.org/10.1111/1755-0998.12682.

David, M., Dursi, L.J., Yao, D., Boutros, P.C., Simpson, J.T., 2017. Nanocall: an open source basecaller for Oxford Nanopore sequencing data. Bioinformatics 33, 49–55.

Eckert, S.E., Chan, J.Z.-M., Houniet, D., PATHSEEK consortium, Breuer, J., Speight, G., 2016. Enrichment by hybridisation of long DNA fragments for nanopore sequencing. Microb. Genom. 2, e000087.

Garalde, D.R., Snell, E.A., Jachimowicz, D., Heron, A.J., Bruce, M., Lloyd, J., Warland, A., Pantic, N., Admassu, T., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, J., Jayasinghe, L., Wright, C., Blasco, J., Sipos, B., Young, S., Juul, S., Clarke, J., Turner, D.J., 2016. Highly, parallel direct RNA sequencing on an array of nanopores. bioRxiv 068809.

Greninger, A.L., Naccache, S.N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J.M., Dodd, R., Mulembakani, P., Schneider, B.S., Muyembe-Tamfum, J.-J., Stramer, S.L., Chiu, C.Y., 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7, 99.

Hall, R., Prow, N., Pyke, A., 2011. Ross River Virus, In: Molecular Detection of Human Viral Pathogens. CRC Press Boca Raton, FL, USA, pp. 349–359.

Hoenen, T., Groseth, A., Rosenke, K., Fischer, R.J., Hoenen, A., Judson, S.D., Martellaro, C., Falzarano, D., Marzi, A., Squires, R.B., Wollenberg, K.R., de Wit, E., Prescott, J., Safronetz, D., van Doremalen, N., Bushmaker, T., Feldmann, F., McNally, K., Bolay, F.K., Fields, B., Sealy, T., Rayfield, M., Nichol, S.T., Zoon, K.C., Massaquoi, M., Munster, V.J., Feldmann, H., 2016. Nanopore sequencing as a rapidly deployable ebola outbreak tool. Emerg. Infect. Dis. 22, 331–334.

Houldcroft, C.J., Beale, M.A., Breuer, J., 2017. Clinical and biological insights from viral genome sequencing. Nat. Rev. Micro. 15, 183–192.

Jansen, H.J., Liem, M., Jong-Raadsen, S.A., Dufour, S., Weltzien, F.-A., Swinkels, W., Koelewijn, A., Palstra, A.P., Pelster, B., Spaink, H.P., Thillart, G.E.V., den Dirks, R.P., Henkel, C.V., 2017. Rapid de novo assembly of the European eel genome from nanopore sequencing reads. Sci. Rep. 7, 7213.

Jaworski, E., Routh, A., 2017. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. PLoS Pathog. 13, e1006365.

Jones, A., Lowry, K., Aaskov, J., Holmes, E.C., Kitchen, A., 2010. Molecular evolutionary dynamics of Ross River virus and implications for vaccine efficacy. J. Gen. Virol. 91, 182–188.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kilianski, A., Haas, J.L., Corriveau, E.J., Liem, A.T., Willis, K.L., Kadavy, D.R., Rosenzweig, C.N., Minot, S.S., 2015. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. GigaScience 4, 12.

Kilianski, A., Roth, P.A., Liem, A.T., Hill, J.M., Willis, K.L., Rossmaier, R.D., Marinich, A.V., Maughan, M.N., Karavis, M.A., Kuhn, J.H., Honko, A.N., Rosenzweig, C.N., 2016. Use of unamplified RNA/cDNA-Hybrid nanopore sequencing for rapid detection and characterization of RNA viruses. Emerg. Infect. Dis. 22, 1448–1451.

Knope, K.E., Muller, M., Kurucz, N., Doggett, S.L., Feldman, R., Johansen, C.A., Hobby, M., Bennett, S., Lynch, S., Sly, A., Currie, B.J., 2016. The national arbovirus and malaria advisory committee, 2016. arboviral diseases and malaria in Australia, 2013-14: annual report of the national arbovirus and malaria advisory committee. Commun. Dis. Intell. 40, E401–E436.

Krueger, F., 2015. Trim Galore: A Wrapper Tool Around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files [WWW Document]. URL http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/(Accessed 12.20.17).

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio].

Lindsay, M.D.A., Jardine, A., Giele, C., Armstrong, P., McCarthy, S., Whittle, A., Pal, N., Lyttle, H., Harrington, S., Nicholson, J., Smith, D., 2015. Investigation of the first case of dengue virus infection acquired in western Australia in seven decades: evidence of importation of infected mosquitoes? PLoS Negl. Trop. Dis. 9, e0004114.

Loman, N.J., Quinlan, A.R., 2014. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics 30, 3399–3401.

Loman, N.J., Quick, J., Simpson, J.T., 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Meth. 12, 733–735.

Lu, H., Giordano, F., Ning, Z., 2016. Oxford nanopore MinION sequencing and genome assembly. Genom. Proteom. Bioinform. 14, 265–279.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17, 10–12.

Medlock, J.M., Hansford, K.M., Schaffner, F., Versteirt, V., Hendrickx, G., Zeller, H., Van Bortel, W., 2012. A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. Vector Borne Zoonotic Dis. 12, 435–447.

Mitsuhashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J.S., Shiraishi, Y., Asano, K., Imanishi, T., 2017. A portable system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer. Sci. Rep. 7, 5657.

Mongan, A.E., Yusuf, I., Wahid, I., Hatta, M., 2015. The evaluation on molecular techniques of reverse transcription loop-Mediated isothermal amplification (RT-LAMP), reverse transcription polymerase chain reaction (RT-PCR), and their diagnostic results on MinION nanopore sequencer for the detection of dengue virus serotypes. Am. J. Microbiol. Res. 3, 118–124.

Patterson, J., Sammon, M., Garg, M., 2016. Dengue, zika and chikungunya: emerging arboviruses in the new world. West J. Emerg. Med. 17, 671–679.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J.H.J., Becker-Ziaja, B., Boettcher, J.P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L.L., Doerrbecker, J., Enkirch, T., García-Dorival, I., Hetzelt, N., Hinzmann, J., Holm, T., Kafetzopoulou, L.E., Koropogui, M., Kosgey, A., Kuisma, E., Logue, C.H., Mazzarelli, A., Meisel, S., Mertens, M., Michel, J., Ngabo, D., Nitzsche, K., Pallasch, E., Patrono, L.V., Portmann, J., Repits, J.G., Rickett, N.Y., Sachse, A., Singethan, K., Vitoriano, I., Yemanaberhan, R.L., Zekeng, E.G., Racine, T., Bello, A., Sall, A.A., Faye, O., Faye, O., Magassouba, N., Williams, C.V., Amburgey, V., Winona, L., Davis, E., Gerlach, J., Washington, F., Monteil, V., Jourdain, M., Bererd, M., Camara, A., Somlare, H., Camara, A., Gerard, M., Bado, G., Baillet, B., Delaune, D., Nebie, K.Y., Diarra, A., Savane, Y., Pallawo, R.B., Gutierrez, G.J., Milhano, N., Roger, I., Williams, C.J., Yattara, F., Lewandowski, K., Taylor, J., Rachwal, P., Turner, D.J., Pollakis, G., Hiscox, J.A., Matthews, D.A., O'Shea, M.K., Johnston, A.M., Wilson, D., Hutley, E., Smit, E., Di Caro, A., Wölfel, R., Stoecker, K., Fleischmann, E., Gabriel, M., Weller, S.A., Koivogui, L., Diallo, B., Keïta, S., Rambaut, A., Formenty, P., et al., 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature 530, 228–232.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., Burton, D.R., Lewis-Ximenez, L.L., Jesus, J.G., de Giovanetti, M., Hill, S., Black, A., Bedford, T., Carroll, M.W., Nunes, M., Alcantara, L.C., Sabino, E.C., Baylis, S.A., Faria, N., Loose, M., Simpson, J.T., Pybus, O.G., Andersen, K.G., Loman, N.J., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat. Protoc. 12, 1261–1276.

Temmam, S., Davoust, B., Berenger, J.-M., Raoult, D., Desnues, C., 2014. Viral metagenomics on animals as a tool for the detection of zoonoses prior to human infection? Int. J. Mol. Sci. 15, 10377–10397.

Tyson, J., O'Neil, N., Jain, M., Olsen, H., Hieter, P., Snutch, T., 2017. Whole genome sequencing and assembly of a Caenorhabditis elegans genome with complex genomic rearrangements using the MinION sequencing device. bioRxiv 099143.

Vaux, A.G.C., Medlock, J.M., 2015. Current status of invasive mosquito surveillance in the UK. Parasites Vectors 8, 351.

Walter, M.C., Zwirglmaier, K., Vette, P., Holowachuk, S.A., Stoecker, K., Genzel, G.H., Antwerpen, M.H., 2016. MinION as part of a biomedical rapidly deployable laboratory. J. Biotechnol. 250, 16–22.

Wang, J., Moore, N.E., Deng, Y.-M., Eccles, D.A., Hall, R.J., 2015. MinION nanopore sequencing of an influenza genome. Front. Microbiol. 6, 766.

Zhong, D., Lo, E., Hu, R., Metzger, M.E., Cummings, R., Bonizzoni, M., Fujioka, K.K., Sorvillo, T.E., Kluh, S., Healy, S.P., Fredregill, C., Kramer, V.L., Chen, X., Yan, G., 2013. Genetic analysis of invasive *Aedes albopictus* populations in Los Angeles County, California and its potential public health impact. PLoS One 8, e68586.

CHAPTER 6

General discussion

6.1. Overview of the research

Monitoring arboviral activity in mosquito populations is essential for the prevention and control of arboviral outbreaks in human and animal populations. The techniques traditionally used by surveillance programs to screen mosquitoes for arboviruses are either low-throughput or require *a priori* knowledge of the virus, which limits surveillance capacity and range of detection. The utilisation of a high-throughput, untargeted approach, such as metatranscriptomic sequencing, would enable upscaled surveillance of mosquito populations, capable of detecting both known and novel viruses, providing valuable genomic and phylogenetic information, and identifying mosquitoes to species level, in a time and cost-efficient way. The incorporation of metatranscriptomics into arbovirus surveillance programs does, however, require the development of mosquito-specific protocols with a clear understanding of the sensitivity and specificity of arbovirus detection relative to current surveillance methods. Furthermore, metatranscriptomic data provides a comprehensive profile of the microbial and invertebrate community within a trap that could be further explored to expand the scope of existing surveillance programs.

The aim of this PhD project was to investigate how mosquito-based arbovirus surveillance can be enhanced using metatranscriptomic sequencing. The discussion in this chapter addresses the following questions:

- How can metatranscriptomics be optimised to detect arboviruses from pools of mosquitoes?
- What is the analytical sensitivity and specificity of metatranscriptomic arbovirus detection?
- Does metatranscriptomics enhance arbovirus surveillance when using unsorted, bulk mosquito traps collected from a variety of field locations over time?
- How does metatranscriptomic sequencing compare to existing surveillance methods and is it sensitive enough for public health outcomes?

- Other than arbovirus detection, what other ways can metatranscriptomic data be used to enhance mosquito-based arbovirus surveillance?
- Is long-read nanopore sequencing suitable for metatranscriptomic arbovirus detection?

6.2. Summary of key findings

6.2.1. Development of resources enabling the use of metatranscriptomics in routine arbovirus surveillance

A number of essential resources enabling the use of metatranscriptomics in routine arbovirus surveillance have been developed as part of this PhD project, including an RNA-seq protocol that enables simultaneous detection of arboviruses and identification of mosquito and biting midge species from unsorted, bulk traps; mosquito-specific ribosomal RNA (rRNA) depletion probes that increase the sensitivity of metatranscriptomic arbovirus detection; curated reference sequence databases for use during bioinformatics analysis; and a criterion for positive detection of arboviruses from metatranscriptomic data based on sensitivity and specificity measurements from a virus spiking study.

The RNA-Seq protocol that was developed (Chapter 2, Batovska et al., 2019) can be used to simultaneously detect arboviruses and identify mosquito and biting midge species from unsorted, bulk traps (Chapter 3). Many surveillance programs identify and sort mosquitoes to species level prior to screening for arboviruses, causing a major bottleneck (Akıner et al., 2019; Knope et al., 2019; Ochieng et al., 2013; Oliver et al., 2018). By simultaneously identifying mosquitoes to species-level and detecting arboviruses in large, unsorted samples, screening can be dramatically upscaled, thereby increasing the speed and capacity of surveillance programs while reducing dependency on experts, such as entomologists, cell culture teams, and molecular biologists (Martinez et al., 2020). Upscaling the number of mosquitoes that are screened increases the chance of detecting circulating arboviruses in the community (Gu and Novak, 2004). Previous studies have used unsorted mosquito samples (Frey et al., 2016; Sadeghi et al., 2018; Shi et al., 2016); however, few have used large (i.e. 1,000 mosquitoes or more) unsorted samples as input for sequencing (Xiao et al., 2018).

The RNA-Seq protocol was further optimised for mosquito samples with the design of mosquito-specific rRNA depletion probes for use during library preparation (Chapter 2, Batovska et al., 2019). While mosquito-specific rRNA depletion has been achieved before (Fauver et al., 2019; Kukutla et al., 2013), it is important to adapt the approach to the species being sequenced due to the specificity of the probes used. The advantage of the chosen method is that rRNA depletion probes can easily be added or removed to customise the protocol as needed. The mosquito rRNA sequences used for probe design are publicly available (https://doi.org/10.6084/m9.figshare.9491258.v1) and could be customised by surveillance programs to include different mosquito species, or other insects that might be found in traps, such as biting midges or moths.

In addition to the rRNA depletion probe sequences, other publicly available databases established as part of this PhD project include the Australian arbovirus sequence database used for targeted arbovirus screening (https://doi.org/10.6084/m9.figshare.12055830.v1) and the cytochrome oxidase I (COI) database used for determining trap mosquito and biting midge species composition (https://doi.org/10.6084/m9.figshare.10246826.v3). While these databases are specific to Australia and south-eastern Australia respectively, they provide a framework that can be expanded, thereby broadening viral and insect detection. These databases can be used by surveillance programs to produce consistent and comparable results across sites and over time. Curated databases are important as they help ensure the reference sequences used for bioinformatic analysis are relevant, accurate, up to date, and as complete as possible (Méric et al., 2019).

A positive arbovirus detection criterion from metatranscriptomic data was also established (Chapter 2, Batovska et al., 2019) and refined (Chapter 3) as part of this PhD project. The criterion put forward defines a positive detection as an arbovirus with a Percent Coverage by Reads ratio (%CR-r) ≥ 2. The ratio is calculated by taking the percent coverage of an arbovirus genome by metatranscriptomic reads and dividing it by any corresponding coverage in the negative controls. By normalising the coverage using the negative control, any contamination or index cross-talk that may have occurred during the preparation and sequencing of the sample can be accounted for. Positive arbovirus detection criteria are important for consistent interpretation of metatranscriptomic sequencing data for routine surveillance activities (Schlaberg et al., 2017).

6.2.2. Sensitivity and specificity of metatranscriptomic arbovirus detection

A spiking experiment was used to demonstrate metatranscriptomic arbovirus detection from mosquito samples is both highly sensitive and highly specific (Chapter 2, Batovska et al., 2019). The experiment involved spiking different dilutions of Ross River virus (RRV), a single-stranded, positive-sense RNA virus, and Umatilla virus, a double-stranded RNA virus, into subsamples of a pool of 100 mosquitoes. The 1:1 dilution represented a biologically relevant viral load of a single experimentally RRV-infected mosquito in a pool of 100 mosquitoes. The full-length genome of both RRV and UMAV could be detected in subsamples spiked with a 1:400 dilution of each virus, suggesting that metatranscriptomic sequencing is sensitive enough to detect one positive mosquito in 40,000. This level of sensitivity permits large-scale arbovirus surveillance that is capable of handling large influxes of mosquito numbers as a result of climatic events that promote breeding.

As a result of the spiking experiment, an initial criterion was developed based on the percentage of arbovirus genome coverage by contigs (%CC) and also the average fold coverage by reads (FCR) (Chapter 2, Batovska et al., 2019). However, when this criterion was applied to metatranscriptomic data generated from unsorted, bulk mosquito traps, it was found to be less sensitive compared to a criterion based on percentage of arbovirus genome coverage by reads (%CR) (Chapter 3). This incongruency was a result of index-cross talk caused by the high viral titres used in the spiking study, which affected the thresholds used for positive detection. When the field traps were sequenced, a blocking reagent was used to reduce index cross-talk, so the previously established thresholds were no longer suitable and required reassessment. Future metatranscriptomic experiments using field-collected traps will help to refine and further validate the criteria used for arbovirus detection and the associated thresholds, which should be regularly assessed to ensure any changes in the samples or protocol have not affected the sensitivity and specificity of detection (Schlaberg et al., 2017).

6.2.3. Enhancing arbovirus surveillance with metatranscriptomics

When applied to unsorted, bulk mosquito traps collected from a variety of locations over a seven-week period by the Victorian Arbovirus Disease Control Program (VADCP), metatranscriptomics enabled the detection of five arboviruses of relevance to public

health (Chapter 3). Whole genome sequences were acquired for most of these arboviruses, providing valuable information for surveillance. Phylogenetic analysis using whole genome sequences is at the core of genomic epidemiology, which can help reveal information about an arbovirus outbreak, such as where the outbreak originated, how it is spreading, the emergence of highly transmissible strains, host range of different lineages, occurrence of viral recombination, and mitigatable predictors of viral dispersal (Pollett et al., 2020). Genomic epidemiology has been used to combat arboviral outbreaks such as yellow fever (Faria et al., 2018), chikungunya (Nunes et al., 2015) and Zika (Grubaugh et al., 2019), with its utility currently exemplified by the SARS-CoV-2 outbreak, the novel coronavirus responsible for the present-day global COVID-19 pandemic (Fauver et al., 2020; Lu et al., 2020). The ability to capture whole genome information is vital to effective outbreak response and needs to be incorporated into arbovirus surveillance programs now to prevent delays resulting from the reactive establishment of laboratory and computational capacity.

Apart from arbovirus detection, the metatranscriptomic data was also used to determine the insect species composition of the unsorted, field-collected traps, enabling the detection of both mosquito and biting midge species (Chapter 3). The simultaneous detection of arboviruses and vector species is a clear demonstration of how metatranscriptomic sequencing can enhance surveillance activities. It also expands surveillance output, as the VADCP currently does not record the presence of biting midge species in traps, despite their ability to transmit viruses of veterinary importance such as bluetongue virus and bovine ephemeral fever virus (Mellor et al., 2000). One disadvantage of using unsorted samples is the inability to reliably link detected arboviruses to their mosquito vector. This is also a problem faced by other arbovirus surveillance approaches such as sentinel animals or the testing of mosquito excreta or saliva via FTA cards (van den Hurk et al., 2012). However, arbovirus vector associations are well-established (Mackenzie et al., 1994; Russell and Dwyer, 2000) and arbovirus detections sans vector identification is better than no testing at all due to limited resources.

6.2.4. Additional outcomes relevant to public health

In addition to the five medically important arboviruses in south-eastern Australia, analysis of the metatranscriptomic trap data revealed a plethora of other viruses in the field-collected mosquitoes (Chapter 3). Further investigation indicated these viruses were likely insect-specific viruses (ISVs), reflecting the findings of other mosquito virome studies (Öhlund et al., 2019; Sadeghi et al., 2018; Shi et al., 2017). Comparison to existing viruses indicated that some of the detections were divergent and likely represent new viruses. Prior studies have used metatranscriptomic sequencing to characterise viral diversity in mosquitoes from Western Australia (Shi et al., 2017) and eastern Australia (Colmant et al., 2017); however, this is the first time the viral diversity of mosquitoes from south-eastern Australia has been metatranscriptomically explored. With over 50,000 mosquitoes sequenced, this is also the largest mosquito metatranscriptomic study to date. Characterisation of the virome endemic to mosquito populations is necessary to establish a baseline for diversity, which makes it possible to detect new or emerging viruses that are of importance to public health, thereby bolstering surveillance. The discovery of new ISVs is also useful for investigations into novel biocontrol agents, as some ISVs have been shown to reduce vector transmission of arboviruses of public health importance (Goenaga et al., 2015; Hall-Mendelin et al., 2016; Romo et al., 2018).

While the majority of the divergent viruses detected in the metatranscriptomic trap data belonged to insect-specific viral families or genera, one detection belonged to the medically important *Alphavirus* genus (Chapter 4, Batovska et al., 2020). The *Alphavirus* genus contains zoonotic viral species such as *Chikungunya virus* and *Eastern equine encephalitis virus*, however, the detected virus grouped with a recently formed clade containing ISVs such as *Eilat virus* (Nasar et al., 2012). The full genome of the virus was published (GenBank accession MN733821) and it was named Yada Yada virus (YYV), the first insect-specific alphavirus in the Asia-Pacific region. Although YYV does not have a direct effect on public health outcomes, its close relationship with pathogenic alphaviruses means it can be used to better understand the evolution of host switching (Li et al., 2015) and has potential applications in diagnostics and vaccine development (Hobson-Peters et al., 2019). Further characterisation of YYV will require isolation of the virus in cell culture, which has been achieved for only two of the other four insect-specific alphaviruses (Hermanns et al., 2020; Nasar et al., 2012). Issues isolating these viruses likely

stem from their narrow mosquito vector range and may require genus-specific cell lines (Torii et al., 2018). Due to the lack of known vector species for YYV, a cDNA clone generated using reverse genetics may be the best approach (Nasar et al., 2012).

6.2.5. Comparison of metatranscriptomics to established surveillance methods

The performance of metatranscriptomic arbovirus detection was compared to reverse transcription quantitative PCR (RT-qPCR) and droplet digital PCR (RT-ddPCR) (Chapter 2, Batovska et al., 2019). Furthermore, RT-qPCR was used to confirm metatranscriptomic arbovirus detections in the field-collected traps (Chapter 3). In Chapter 2, RT-qPCR and RT-ddPCR both detected the lowest concentration RRV and UMAV spike subsamples (1:160,000), whereas metatranscriptomic sequencing detected the 1:400 spike subsamples and above. In Chapter 3, there were a total of 43 arbovirus detections made using RT-qPCR, whereas there were 33 arbovirus detections made using metatranscriptomic sequencing. Previous studies have corroborated the higher sensitivity of PCR-based methods when detecting viruses from complex input material containing high levels of host background (Bibby and Peccia, 2013; Fernandez-Cassi et al., 2017, 2018). However, the ability to detect minute levels of virus may not necessarily confer the most effective surveillance results. Arboviruses must sufficiently replicate in order to successfully disseminate throughout a mosquito and be transmitted in their saliva, suggesting that higher viral loads are more indicative of active local transmission in the community (Vazeille et al., 2019). With the ability to detect a single RRV-infected mosquito in a pool of 40,000 mosquitoes, metatranscriptomic sequencing is capable of detecting viral loads relevant to public health outcomes. Furthermore, metatranscriptomics allows untargeted virus detection, unlike PCR-based methods that require knowledge of the viral sequence. In Chapter 3, metatranscriptomics was used to screen mosquitoes for 74 known arboviruses of medical importance and explore the broader virome, leading to the detection of five arboviruses and dozens of known and novel ISVs, many with full genome sequences to allow genomic epidemiological analysis. Additionally, the mosquito and biting midge species in the trap were also identified using the metatranscriptomic data. To be able to screen for and detect the same level of information using PCR-based methods would be time, labour, and cost-prohibitive.

When the detections of medically important arboviruses and mosquito species in the same field-collected traps were considered, the results of the metatranscriptomic and VADCP surveillance approaches were similar (Chapter 3). Four out of five arboviruses were detected by the cell culture-based VADCP methods, with UMAV detected only via metatranscriptomics. As for mosquito species, the morphology-based VADCP methods detected 8 out of the 12 species detected metatranscriptomically. Some of the differences observed may be attributed to the separate subsamples used per trap for each surveillance approach, while others are a result of the differing sensitivities of the methods used. For instance, the sensitivity of morphological identification is lowered when mosquitoes share similar morphology or have been damaged by the field trap (Cansado-Utrilla et al., 2020; Rodrigues et al., 2014). The overall similarity of the results confirms the ability of metatranscriptomics to achieve the surveillance outcomes currently achieved by traditional screening methods. Additionally, metatranscriptomics offers many advantages, including the ability to considerably upscale the number of mosquitoes that can be screened simultaneously, reduce the time and expertise required to perform the surveillance, and expand the range of detection, both for viral and vector species.

The cost of metatranscriptomic arbovirus detection and mosquito species identification was approximately AUD$230 per sample containing up to 1,000 mosquitoes. Protocol alterations, such as in-house rRNA depletion and using the Illumina NovaSeq instead of the HiSeq, could potentially reduce this to AUD$110 per sample (Batovska et al., 2019). In comparison, PCR for four arboviral targets per trap costs approximately AUD$100, and morphological mosquito species identification costs an additional AUD$120 for a trap containing 1,000 mosquitoes (Lynch et al., 2020). Based on arbovirus detection and vector species identification alone, metatranscriptomics is a more cost-effective surveillance tool than PCR for screening traps due to its ability to detect an essentially unlimited number of targets.

6.2.6. Long-read sequencing for metatranscriptomic arbovirus detection

In addition to short-read sequencing, this PhD project also included an investigation into the use of long-read sequencing for metatranscriptomic arbovirus surveillance. The

Oxford Nanopore Technologies (ONT) MinION sequencer was used to detect RRV from an infected mosquito using a metatranscriptomic approach (Chapter 5, Batovska et al., 2017). This proof-of-concept study was the first time the MinION sequencer was used for metatranscriptomic arbovirus detection from a mosquito sample, with later studies demonstrating this using pools of mosquitoes (Russell et al., 2018). When compared to short-read sequencing on the MiSeq, the MinION generated poor quality reads; however, RRV consensus sequences were of comparable quality for both platforms. Long reads are beneficial for surveillance as they provide more genetic and epidemiological information on viral strains (Russell et al., 2018), resolve viral haplotypes (Quick et al., 2016), and can recover entire viral genomes within a single read, surpassing the need for read assembly (Beaulaurier et al., 2020). The MinION has unique features suited to surveillance, specifically its small, portable size enabling in-field sequencing, its low instrument cost permitting smaller surveillance programs to invest in sequencing, and its real-time sequencing, which allows detection of arboviruses as they are being sequenced (Gardy et al., 2015). These advantages, along with the continuously improving error rate and throughput, warrant further investigation of long-read nanopore sequencing for metatranscriptomic arbovirus surveillance.

6.3. Future directions and recommendations

6.3.1. Protocols and resources used for metatranscriptomic sequencing

Further optimisation of the metatranscriptomic arbovirus detection protocol developed as part of this PhD project would increase assay sensitivity and make it more cost-effective when compared to current surveillance methods. One of the first steps of the protocol involves extracting nucleic acid from homogenised mosquitoes using a column-based kit (Chapter 2, Batovska et al., 2019). A recent study showed that using magnetic beads to extract nucleic acid from homogenised insects can increase viral reads during metatranscriptomic sequencing by 120.8 fold, which could possibly be attributed to the lack of clogging by cell debris that can occur when using columns (Akello et al., 2020). Further increases in viral reads could be achieved by expanding the range of rRNA probes used for depletion during library preparation, ensuring that they are specific to the species

being trapped during surveillance. In-house alternatives to the kit-based depletion used in the protocol should be explored to lower costs (Culviner et al., 2020; Gu et al., 2016). Viral enrichment could also increase the sensitivity and specificity of metatranscriptomic sequencing, with options such as pan-viral amplification (Deng et al., 2020) and probe capture (Briese et al., 2015; Metsky et al., 2019) helping to maximise viral reads while still allowing a broad range of detection.

Apart from increasing the proportion of viral reads, cost per sample could also be decreased by multiplexing large numbers of trap samples on ultra-high-throughput sequencers, such as the Illumina NovaSeq. When multiplexing, unique dual indexing should be used to reduce index cross-talk that can lead to false positive results (Costello et al., 2018; MacConaill et al., 2018). Currently, commercially available options are limited to ≤384 unique dual indexes and so in-house index design and validation may be required for larger surveillance activities (Glenn et al., 2019). High levels of multiplexing would also require automation of the library preparation steps, which would also decrease the time and labour involved for sample preparation (Chiu and Miller, 2019). While increased sequencing throughput can lower costs, the large amount of data generated will need to be stored so long-term data management should be taken into consideration.

Utilisation of metatranscriptomic sequencing in arbovirus surveillance programs will require the development of user-friendly bioinformatics analysis programs. Currently, metatranscriptomic analysis is performed by staff with bioinformatic expertise using customised pipelines that vary in choice of software, thresholds, parameters, and strategy (Nooij et al., 2018). Many of the initial pipeline steps such as read trimming, removing host reads, *de novo* assembly, and read mapping can be automated. However, the output is often a long list of detected viruses, many of which have not yet been assigned a taxonomic classification. It is difficult to know whether these unclassified viruses are capable of infecting vertebrates without further investigation into the history and phylogeny of each virus, which creates an analysis bottleneck and delays reportable outcomes for surveillance (Simmonds, 2015). Future efforts should focus on creating a bioinformatic tool that automates the detection of vertebrate viruses from assembled metatranscriptomic data and can flag the presence of unclassified or novel viruses with a higher risk of infecting vertebrates that may need to be further investigated.

Development of a metatranscriptomic protocol that preserves the infectivity of viral particles would further enhance arbovirus surveillance by enabling more in-depth investigation of the novel viruses detected via sequencing. The majority of methods used to extract nucleic acid from homogenised mosquitoes result in the inactivation of viruses, preventing downstream applications that require a viral isolate (Temmam et al., 2015). These applications include competence testing to determine the virus host range and vector specificity, pathogenesis testing to assess disease potential, and serological testing to investigate antigenicity (Brault and Blitvich, 2018). A metatranscriptomic protocol that preserves viral infectivity can be achieved by altering the viral purification steps and reducing host DNA and RNA contaminants (Temmam et al., 2015). Another option is to store an aliquot of the mosquito trap homogenates in conditions that retain virus infectivity, so that if sequencing identifies a virus of interest, it can be isolated and characterised using the original mosquito sample (De Paoli, 2005). By maintaining the ability to isolate viruses for further characterisation, the findings of metatranscriptomic arbovirus surveillance can be fully understood and used to inform control and prevention strategies.

6.3.2. Standardisation of metatranscriptomic protocols

Routine surveillance requires robust, standardised protocols that have been validated to produce reliable and reproducible results. Metatranscriptomics is a relatively new field, and as such the reagents, sequencers, and bioinformatic tools available are constantly changing without any definitive guidelines or requirements on how to validate metatranscriptomic protocols (Chiu and Miller, 2019). Any changes can affect detection sensitivity and specificity, and therefore it is necessary to develop a method to regularly validate the metatranscriptomic protocol in response to any modifications, which might lead to updating the sequencing depth, limits of detection, or detection thresholds (Gu et al., 2019). The untargeted nature of metatranscriptomics creates a unique challenge as it is difficult to validate the detection of an essentially unlimited number of organisms; however, custom mixtures containing a diverse range of microbes or synthetic RNA can be used to regularly test performance metrics and ensure assay accuracy (Hardwick et al., 2018). While a range of these are commercially available (Mason et al., 2017), most are

not designed for metatranscriptomic virus detection, representing a future avenue for development.

6.3.3. Further enhancements of mosquito-based arbovirus surveillance using metatranscriptomics

The versatility of metatranscriptomic sequencing means that it can be applied to many different sample types, creating opportunities to further enhance arbovirus surveillance. For instance, instead of metatranscriptomically screening mosquitoes trapped via CO_2 light traps, mosquitoes caught using gravid traps could be screened instead. Since gravid mosquitoes are typically blood-fed, they are more likely to contain arboviruses, and can also provide information on mosquito host-feeding patterns (Tomazatos et al., 2019). Additionally, blood meals can be used to conduct xenosurveillance, where mosquitoes serve as "flying syringes" to survey the environment for the circulation of host-specific, non-arboviruses, such as hepatitis B in humans (Fauver et al., 2018), canine distemper virus in dogs (Grubaugh et al., 2015), and avian influenza virus in poultry (Barbazan et al., 2008). Other sample types include mosquito excreta collected from the bottom of traps (Ramírez et al., 2018), and mosquito saliva deposited on FTA cards (Birnberg et al., 2020), the latter of which has the advantage of detecting only arboviruses that are transmitted by the mosquito via feeding. All of these alternative sampling approaches could be used to perform more targeted metatranscriptomics to achieve specific public health outcomes.

Apart from virus detection and mosquito species identification, metatranscriptomic trap data can be mined for additional surveillance-relevant information. Arbovirus detections could be further explored by performing variant analyses to quantify and characterise genetic diversity of virus subpopulations present within traps (Grubaugh et al., 2017). The screening of metatranscriptomic data could be expanded beyond arboviruses to the detection of bacteria (Ramos-Nino et al., 2020), fungi (Chandler et al., 2015), parasites (Shi et al., 2017), and endosymbionts such as *Wolbachia* (Hall-Mendelin et al., 2013). Due to the large amount of mosquito sequences in metatranscriptomic data, it could also be used for gene annotation of mosquito genome assemblies (Prasad et al., 2017). The mosquito metatranscriptomic data could also be screened for the presence of variants in genes

associated with insecticide resistance, such as the sodium channel gene (Endersby-Harshman et al., 2020) and cytochrome P450 (Faucon et al., 2015). The COI-based insect species identification could also be further developed to find divergent COI sequences for the discovery of novel species, or the generation of reference material for known species that have not yet been barcoded. These examples of how the one metatranscriptomic dataset can be utilised highlights the wide-ranging applicability of the technique, which will be further expanded with the development of new computational methods.

While this PhD project has focused on enhancing arbovirus surveillance for public health outcomes, metatranscriptomic sequencing could be used to help answer broader ecological questions. For instance, the large diversity of mosquito viruses found in the metatranscriptomic data could be further explored to gain a better understanding of their emergence and evolution (Li et al., 2015). Although many studies have explored the viral diversity of mosquito populations, few have investigated how the mosquito virome is transmitted across different developmental stages or how it changes over time. Gaining a better understanding of the core mosquito virome can lead to insights into how it influences vector competence and the transmission of arboviruses relevant to public health (Agboli et al., 2019; Patterson et al., 2020). Metatranscriptomics could also be used to characterise the virome of parasitic mites that are occasionally found on mosquitoes (Simmons and Hutchinson, 2016), and investigate whether they play a role in arbovirus transmission.

Now that metatranscriptomic arbovirus detection from mosquito samples has been demonstrated on the portable MinION sequencer (Batovska et al., 2017; Russell et al., 2018), further investigation is required to develop an in-field protocol for mobile surveillance activities. Similar to short-read sequencing, the sensitivity and specificity of long-read nanopore sequencing needs to be established for use with bulk, unsorted mosquito pools. A mobile surveillance protocol will require cost-effective and rapid in-field sample preparation, multiplexing of samples on the MinION, and remote data analysis capable of basecalling nanopore sequencing signals (Gowers et al., 2019). The recently released Flongle, a small, single-use flow cell for the MinION, and VolTRAX II, a portable device that automates library preparation, could be used to build in-field capabilities (Hall et al., 2020). Other advancements in nanopore technology, such as direct

RNA sequencing, should also be investigated to assess their ability to enhance arbovirus surveillance.

6.4. Concluding remarks

The work presented in this thesis demonstrates how metatranscriptomics can be used to revolutionise arbovirus surveillance. The methods and resources that were developed provide a comprehensive framework for the use of metatranscriptomic sequencing to support an active surveillance program for enhanced detection of arboviruses of public health significance. Further work is required to define the experimental and validation parameters of metatranscriptomic sequencing and ensure that the reported output is clearly defined and easily interpreted. Once established, automated high-throughput sequencing workflows and bioinformatics pipelines will enable cost-effective metatranscriptomics to be incorporated into routine, mosquito-based arbovirus surveillance for improved public health outcomes.

6.5. References

Agboli, E., Leggewie, M., Altinli, M., and Schnettler, E. (2019). Mosquito-specific viruses - transmission and interaction. Viruses *11*.

Akello, J.O., Leib, S.L., Engler, O., and Beuret, C. (2020). Evaluation of viral RNA recovery methods in vectors by metagenomic sequencing. Viruses *12*, 562.

Akıner, M.M., Öztürk, M., Başer, A.B., Günay, F., Hacıoğlu, S., Brinkmann, A., Emanet, N., Alten, B., Özkul, A., Nitsche, A., et al. (2019). Arboviral screening of invasive *Aedes* species in northeastern Turkey: West Nile virus circulation and detection of insect-only viruses. PLOS Neglected Tropical Diseases *13*.

Barbazan, P., Thitithanyanont, A., Missé, D., Dubot, A., Bosc, P., Luangsri, N., Gonzalez, J.-P., and Kittayapong, P. (2008). Detection of H5N1 avian influenza virus from mosquitoes collected in an infected poultry farm in Thailand. Vector Borne and Zoonotic Diseases *8*, 105–109.

Batovska, J., Lynch, S.E., Rodoni, B.C., Sawbridge, T.I., and Cogan, N.O. (2017). Metagenomic arbovirus detection using MinION nanopore sequencing. Journal of Virological Methods *249*, 79–84.

Batovska, J., Mee, P.T., Lynch, S.E., Sawbridge, T.I., and Rodoni, B.C. (2019). Sensitivity and specificity of metatranscriptomics as an arbovirus surveillance tool. Scientific Reports *9*, 1–13.

Batovska, J., Buchmann, J.P., Holmes, E.C., and Lynch, S.E. (2020). Coding-complete genome sequence of Yada Yada virus, a novel alphavirus detected in Australian mosquitoes. Microbiology Resource Announcements *9*.

Beaulaurier, J., Luo, E., Eppley, J.M., Uyl, P.D., Dai, X., Burger, A., Turner, D.J., Pendleton, M., Juul, S., Harrington, E., et al. (2020). Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. Genome Research gr.251686.119.

Bibby, K., and Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environmental Science & Technology *47*, 1945–1951.

Birnberg, L., Temmam, S., Aranda, C., Correa-Fiz, F., Talavera, S., Bigot, T., Eloit, M., and Busquets, N. (2020). Viromics on honey-baited FTA cards as a new tool for the detection of circulating viruses in mosquitoes. Viruses *12*, 274.

Brault, A.C., and Blitvich, B.J. (2018). Continued need for comprehensive genetic and phenotypic characterization of viruses: benefits of complementing sequence analyses with functional determinations. The American Journal of Tropical Medicine and Hygiene *98*, 1213.

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O.J., and Lipkin, W.I. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio *6*, e01491-01415.

Cansado-Utrilla, C., Jeffries, C.L., Kristan, M., Brugman, V.A., Heard, P., Camara, G., Sylla, M., Beavogui, A.H., Messenger, L.A., Irish, S.R., et al. (2020). An assessment of adult mosquito collection techniques for studying species abundance and diversity in Maferinyah, Guinea. Parasites & Vectors *13*, 150.

Chandler, J.A., Liu, R.M., and Bennett, S.N. (2015). RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. Frontiers in Microbiology *6*, 185.

Chiu, C.Y., and Miller, S.A. (2019). Clinical metagenomics. Nature Reviews Genetics *20*, 341–355.

Colmant, A., Etebari, K., Webb, C., Ritchie, S., Jansen, C., van den Hurk, A., Bielefeldt-Ohmann, H., Hobson-Peters, J., Asgari, S., and Hall, R. (2017). Discovery of new orbiviruses and totivirus from *Anopheles* mosquitoes in eastern Australia. Archives of Virology *162*, 1–6.

Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T., et al. (2018). Characterization and remediation of sample index

swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics *19*, 332.

Culviner, P.H., Guegler, C.K., and Laub, M.T. (2020). A simple, cost-effective, and robust method for rRNA depletion in RNA-sequencing studies. MBio *11*.

De Paoli, P. (2005). Biobanking in microbiology: from sample collection to epidemiology, diagnosis and research. FEMS Microbiology Reviews *29*, 897–910.

Deng, X., Achari, A., Federman, S., Yu, G., Somasekar, S., Bártolo, I., Yagi, S., Mbala-Kingebeni, P., Kapetshi, J., Ahuka-Mundeke, S., et al. (2020). Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. Nature Microbiology *5*, 443–454.

Endersby-Harshman, N.M., Schmidt, T.L., Chung, J., van Rooyen, A., Weeks, A.R., and Hoffmann, A.A. (2020). Heterogeneous genetic invasions of three insecticide resistance mutations in Indo-Pacific populations of *Aedes aegypti* (L.). Molecular Ecology *29*, 1628–1641.

Faria, N.R., Kraemer, M.U.G., Hill, S.C., Goes de Jesus, J., Aguiar, R.S., Iani, F.C.M., Xavier, J., Quick, J., du Plessis, L., Dellicour, S., et al. (2018). Genomic and epidemiological monitoring of yellow fever virus transmission potential. Science *361*, 894–899.

Faucon, F., Dusfour, I., Gaude, T., Navratil, V., Boyer, F., Chandre, F., Sirisopa, P., Thanispong, K., Juntarajumnong, W., Poupardin, R., et al. (2015). Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. Genome Research *25*, 1347–1359.

Fauver, J.R., Weger-Lucarelli, J., Iii, L.S.F., Bolay, K., Bolay, F.K., Ii, J.W.D., Brackney, D.E., Foy, B.D., Stenglein, M.D., and Ebel, G.D. (2018). Xenosurveillance reflects traditional sampling techniques for the identification of human pathogens: a comparative study in West Africa. PLOS Neglected Tropical Diseases *12*, e0006348.

Fauver, J.R., Akter, S., Morales, A.I.O., Black, W.C., Rodriguez, A.D., Stenglein, M.D., Ebel, G.D., and Weger-Lucarelli, J. (2019). A reverse-transcription/RNase H based protocol for depletion of mosquito ribosomal RNA facilitates viral intrahost evolution analysis, transcriptomics and pathogen discovery. Virology *528*, 181–197.

Fauver, J.R., Petrone, M.E., Hodcroft, E.B., Shioda, K., Ehrlich, H.Y., Watts, A.G., Vogels, C.B.F., Brito, A.F., Alpert, T., Muyombwe, A., et al. (2020). Coast-to-coast spread of SARS-CoV-2 during the early epidemic in the United States. Cell *181*, 990–996.

Fernandez-Cassi, X., Timoneda, N., Gonzales-Gustavson, E., Abril, J.F., Bofill-Mas, S., and Girones, R. (2017). A metagenomic assessment of viral contamination on fresh parsley plants irrigated with fecally tainted river water. International Journal of Food Microbiology *257*, 80–90.

Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., and Girones, R. (2018). Metagenomics for the

study of viruses in urban sewage as a tool for public health surveillance. The Science of the Total Environment *618*, 870–880.

Frey, K.G., Biser, T., Hamilton, T., Santos, C.J., Pimentel, G., Mokashi, V.P., and Bishop-Lilly, K.A. (2016). Bioinformatic characterization of mosquito viromes within the eastern United States and Puerto Rico: discovery of novel viruses. Evolutionary Bioinformatics Online *12*, 1–12.

Gardy, J., Loman, N.J., and Rambaut, A. (2015). Real-time digital pathogen surveillance - the time is now. Genome Biology *16*, 155.

Glenn, T.C., Nilsen, R.A., Kieran, T.J., Sanders, J.G., Bayona-Vásquez, N.J., Finger, J.W., Pierson, T.W., Bentley, K.E., Hoffberg, S.L., Louha, S., et al. (2019). Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). PeerJ *7*, e7755.

Goenaga, S., Kenney, J.L., Duggal, N.K., Delorey, M., Ebel, G.D., Zhang, B., Levis, S.C., Enria, D.A., and Brault, A.C. (2015). Potential for co-infection of a mosquito-specific flavivirus, Nhumirim virus, to block West Nile virus transmission in mosquitoes. Viruses *7*, 5801–5812.

Gowers, G.-O.F., Vince, O., Charles, J.-H., Klarenberg, I., Ellis, T., and Edwards, A. (2019). Entirely off-grid and solar-powered DNA sequencing of microbial communities during an ice cap traverse expedition. Genes *10*, 902.

Grubaugh, N.D., Sharma, S., Krajacich, B.J., Iii, L.S.F., Bolay, F.K., Ii, J.W.D., Johnson, W.E., Ebel, G.D., Foy, B.D., and Brackney, D.E. (2015). Xenosurveillance: a novel mosquito-based approach for examining the human-pathogen landscape. PLOS Neglected Tropical Diseases *9*, e0003628.

Grubaugh, N.D., Fauver, J.R., Rückert, C., Weger-Lucarelli, J., Garcia-Luna, S., Murrieta, R.A., Gendernalik, A., Smith, D.R., Brackney, D.E., and Ebel, G.D. (2017). Mosquitoes transmit unique West Nile virus populations during each feeding episode. Cell Reports *19*, 709–718.

Grubaugh, N.D., Saraf, S., Gangavarapu, K., Watts, A., Tan, A.L., Oidtman, R.J., Ladner, J.T., Oliveira, G., Matteson, N.L., Kraemer, M.U.G., et al. (2019). Travel surveillance and genomics uncover a hidden Zika outbreak during the waning epidemic. Cell *178*, 1057-1071.e11.

Gu, W., and Novak, R.J. (2004). Short report: detection probability of arbovirus infection in mosquito populations. The American Journal of Tropical Medicine and Hygiene *71*, 636–638.

Gu, W., Crawford, E.D., O'Donovan, B.D., Wilson, M.R., Chow, E.D., Retallack, H., and DeRisi, J.L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biology *17*, 41.

Gu, W., Miller, S., and Chiu, C.Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. Annual Review of Pathology *14*, 319–338.

Hall, C.L., Zascavage, R.R., Sedlazeck, F.J., and Planz, J.V. (2020). Potential applications of nanopore sequencing for forensic analysis. Forensic Science Review *32*, 23–54.

Hall-Mendelin, S., Allcock, R., Kresoje, N., van den Hurk, A.F., and Warrilow, D. (2013). Detection of arboviruses and other micro-organisms in experimentally infected mosquitoes using massively parallel sequencing. PLOS ONE *8*.

Hall-Mendelin, S., McLean, B.J., Bielefeldt-Ohmann, H., Hobson-Peters, J., Hall, R.A., and van den Hurk, A.F. (2016). The insect-specific Palm Creek virus modulates West Nile virus infection in and transmission by Australian mosquitoes. Parasites & Vectors *9*, 414.

Hardwick, S.A., Chen, W.Y., Wong, T., Kanakamedala, B.S., Deveson, I.W., Ongley, S.E., Santini, N.S., Marcellin, E., Smith, M.A., Nielsen, L.K., et al. (2018). Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nature Communications *9*, 1–10.

Hermanns, K., Marklewitz, M., Zirkel, F., Overheul, G.J., Page, R.A., Loaiza, J.R., Drosten, C., van Rij, R.P., and Junglen, S. (2020). Agua Salud alphavirus defines a novel lineage of insect-specific alphaviruses discovered in the New World. The Journal of General Virology *101*.

Hobson-Peters, J., Harrison, J.J., Watterson, D., Hazlewood, J.E., Vet, L.J., Newton, N.D., Warrilow, D., Colmant, A.M.G., Taylor, C., Huang, B., et al. (2019). A recombinant platform for flavivirus vaccines and diagnostics using chimeras of a new insect-specific virus. Science Translational Medicine *11*.

van den Hurk, A.F., Hall-Mendelin, S., Johansen, C.A., Warrilow, D., and Ritchie, S.A. (2012). Evolution of mosquito-based arbovirus surveillance systems in Australia. Journal of Biomedicine & Biotechnology *2012*, 325659.

Knope, K., Doggett, S.L., Jansen, C.C., Johansen, C.A., Kurucz, N., Feldman, R., Lynch, S.E., Hobby, M.P., Sly, A., Jardine, A., et al. (2019). Arboviral diseases and malaria in Australia, 2014–15: Annual report of the National Arbovirus and Malaria Advisory Committee. Communicable Diseases Intelligence (2018) *43*.

Kukutla, P., Steritz, M., and Xu, J. (2013). Depletion of ribosomal RNA for mosquito gut metagenomic RNA-seq. Journal of Visualized Experiments: JoVE *(74)*, e50093.

Li, C.-X., Shi, M., Tian, J.-H., Lin, X.-D., Kang, Y.-J., Chen, L.-J., Qin, X.-C., Xu, J., Holmes, E.C., and Zhang, Y.-Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. ELife *4*.

Lu, J., du Plessis, L., Liu, Z., Hill, V., Kang, M., Lin, H., Sun, J., François, S., Kraemer, M.U.G., Faria, N.R., et al. (2020). Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell *181*, 997-1003.e9.

Lynch, S.E., Mee, P.T., Batovska, J., Brown, K., Crowder, J., and Rodoni, B.C. (2020). Mosquito and arbovirus biodiversity and Ross River virus genomic epidemiology in Victoria, Australia, 2016-2017. *Manuscript in preparation*.

MacConaill, L.E., Burns, R.T., Nag, A., Coleman, H.A., Slevin, M.K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M.S., et al. (2018). Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics *19*, 30.

Mackenzie, J.S., Lindsay, M.D., Coelen, R.J., Broom, A.K., Hall, R.A., and Smith, D.W. (1994). Arboviruses causing human disease in the Australasian zoogeographic region. Archives of Virology *136*, 447–467.

Martinez, B., Reaser, J.K., Dehgan, A., Zamft, B., Baisch, D., McCormick, C., Giordano, A.J., Aicher, R., and Selbe, S. (2020). Technology innovation: advancing capacities for the early detection of and rapid response to invasive species. Biological Invasions *22*, 75–100.

Mason, C.E., Afshinnekoo, E., Tighe, S., Wu, S., and Levy, S. (2017). International standards for genomes, transcriptomes, and metagenomes. Journal of Biomolecular Techniques *28*, 8–18.

Mellor, P.S., Boorman, J., and Baylis, M. (2000). Culicoides biting midges: their role as arbovirus vectors. Annual Review of Entomology *45*, 307–340.

Méric, G., Wick, R.R., Watts, S.C., Holt, K.E., and Inouye, M. (2019). Correcting index databases improves metagenomic studies. BioRxiv 712166.

Metsky, H.C., Siddle, K.J., Gladden-Young, A., Qu, J., Yang, D.K., Brehio, P., Goldfarb, A., Piantadosi, A., Wohl, S., Carter, A., et al. (2019). Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. Nature Biotechnology *37*, 160–168.

Nasar, F., Palacios, G., Gorchakov, R.V., Guzman, H., Rosa, A.P.T.D., Savji, N., Popov, V.L., Sherman, M.B., Lipkin, W.I., Tesh, R.B., et al. (2012). Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. PNAS *109*, 14622–14627.

Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., and Koopmans, M.P.G. (2018). Overview of virus metagenomic classification methods and their biological applications. Frontiers in Microbiology *9*, 749.

Nunes, M.R.T., Faria, N.R., de Vasconcelos, J.M., Golding, N., Kraemer, M.U., de Oliveira, L.F., Azevedo, R. do S. da S., da Silva, D.E.A., da Silva, E.V.P., da Silva, S.P., et al. (2015). Emergence and potential for spread of chikungunya virus in Brazil. BMC Medicine *13*, 102.

Ochieng, C., Lutomiah, J., Makio, A., Koka, H., Chepkorir, E., Yalwala, S., Mutisya, J., Musila, L., Khamadi, S., Richardson, J., et al. (2013). Mosquito-borne arbovirus surveillance at selected sites in diverse ecological zones of Kenya; 2007 – 2012. Virology Journal *10*, 140.

Öhlund, P., Hayer, J., Lundén, H., Hesson, J.C., and Blomström, A.-L. (2019). Viromics reveal a number of novel RNA viruses in Swedish mosquitoes. Viruses *11*, 1027.

Oliver, J., Lukacik, G., Kokas, J., Campbell, S.R., Kramer, L.D., Sherwood, J.A., and Howard, J.J. (2018). Twenty years of surveillance for Eastern equine encephalitis virus in mosquitoes in New York State from 1993 to 2012. Parasites & Vectors *11*, 362.

Patterson, E.I., Villinger, J., Muthoni, J.N., Dobel-Ober, L., and Hughes, G.L. (2020). Exploiting insect-specific viruses as a novel strategy to control vector-borne disease. Current Opinion in Insect Science *39*, 50–56.

Pollett, S., Fauver, J.R., Maljkovic Berry, I., Melendrez, M., Morrison, A., Gillis, L.D., Johansson, M.A., Jarman, R.G., and Grubaugh, N.D. (2020). Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. The Journal of Infectious Diseases *221*, S308–S318.

Prasad, T.S.K., Mohanty, A.K., Kumar, M., Sreenivasamurthy, S.K., Dey, G., Nirujogi, R.S., Pinto, S.M., Madugundu, A.K., Patil, A.H., Advani, J., et al. (2017). Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. Genome Research *27*, 133–144.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A., et al. (2016). Real-time, portable genome sequencing for Ebola surveillance. Nature *530*, 228–232.

Ramírez, A.L., Hall-Mendelin, S., Doggett, S.L., Hewitson, G.R., McMahon, J.L., Ritchie, S.A., and van den Hurk, A.F. (2018). Mosquito excreta: a sample type with many potential applications for the investigation of Ross River virus and West Nile virus ecology. PLOS Neglected Tropical Diseases *12*, e0006771.

Ramos-Nino, M.E., Fitzpatrick, D.M., Eckstrom, K.M., Tighe, S., Hattaway, L.M., Hsueh, A.N., Stone, D.M., Dragon, J.A., and Cheetham, S. (2020). Metagenomic analysis of *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes from Grenada, West Indies. PLOS ONE *15*.

Rodrigues, M.S., Silva, I.M., Leal, L.B., Santos, C.A.C.D., and Eiras, Á.E. (2014). Development of a new mosquito retention system for the BG-malaria trap to reduce the damage to mosquitoes. Journal of the American Mosquito Control Association *30*, 184–190.

Romo, H., Kenney, J.L., Blitvich, B.J., and Brault, A.C. (2018). Restriction of Zika virus infection and transmission in *Aedes aegypti* mediated by an insect-specific flavivirus. Emerging Microbes & Infections *7*, 1–13.

Russell, R.C., and Dwyer, D.E. (2000). Arboviruses associated with human disease in Australia. Microbes and Infection *2*, 1693–1704.

Russell, J.A., Campos, B., Stone, J., Blosser, E.M., Burkett-Cadena, N., and Jacobs, J.L. (2018). Unbiased strain-typing of arbovirus directly from mosquitoes using nanopore sequencing: a field-forward biosurveillance protocol. Scientific Reports *8*, 5417.

Sadeghi, M., Altan, E., Deng, X., Barker, C.M., Fang, Y., Coffey, L.L., and Delwart, E. (2018). Virome of > 12 thousand *Culex* mosquitoes from throughout California. Virology *523*, 74–88.

Schlaberg, R., Chiu, C.Y., Miller, S., Procop, G.W., Weinstock, G., Professional Practice Committee and Committee on Laboratory Practices of the American Society for Microbiology, and Microbiology Resource Committee of the College of American Pathologists (2017). Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Archives of Pathology & Laboratory Medicine *141*, 776–786.

Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S., et al. (2016). Redefining the invertebrate RNA virosphere. Nature *540*, 539–543.

Shi, M., Neville, P., Nicholson, J., Eden, J.-S., Imrie, A., and Holmes, E.C. (2017). High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. Journal of Virology *91*, e00680-17.

Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. The Journal of General Virology *96*, 1193–1206.

Simmons, T.W., and Hutchinson, M.L. (2016). A critical review of all known published records for water mite (Acari: Hydrachnidiae) and mosquito (Diptera: Culicidae) parasitic associations from 1975 to present. Journal of Medical Entomology *53*, 737–752.

Temmam, S., Monteil-Bouchard, S., Robert, C., Pascalis, H., Michelle, C., Jardot, P., Charrel, R., Raoult, D., and Desnues, C. (2015). Host-associated metagenomics: a guide to generating infectious RNA viromes. PLOS ONE *10*, e0139810.

Tomazatos, A., Jansen, S., Pfister, S., Török, E., Maranda, I., Horváth, C., Keresztes, L., Spînu, M., Tannich, E., Jöst, H., et al. (2019). Ecology of West Nile virus in the Danube Delta, Romania: phylogeography, xenosurveillance and mosquito host-feeding patterns. Viruses *11*.

Torii, S., Orba, Y., Hang'ombe, B.M., Mweene, A.S., Wada, Y., Anindita, P.D., Phongphaew, W., Qiu, Y., Kajihara, M., Mori-Kajihara, A., et al. (2018). Discovery of Mwinilunga alphavirus: a novel alphavirus in *Culex* mosquitoes in Zambia. Virus Research *250*, 31–36.

Vazeille, M., Madec, Y., Mousson, L., Bellone, R., Barré-Cardi, H., Sousa, C.A., Jiolle, D., Yébakima, A., de Lamballerie, X., and Failloux, A.-B. (2019). Zika virus threshold determines transmission by European *Aedes albopictus* mosquitoes. Emerging Microbes & Infections *8*, 1668–1678.

Xiao, P., Li, C., Zhang, Y., Han, J., Guo, X., Xie, L., Tian, M., Li, Y., Wang, M., Liu, H., et al. (2018). Metagenomic sequencing from mosquitoes in China reveals a variety of insect and human viruses. Frontiers in Cellular and Infection Microbiology *8*.