Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Caraterização de vírus específicos de insetos: análise da sua diversidade genética e relação com outros vírus

Paulo Jorge Sampaio Morais

DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS BIOMÉDICAS, ESPECIALIDADE EM MICROBIOLOGIA

Julho, 2022

INSTITUTO DE HIGIENE E
MEDICINA TROPICAL
DESDE 1902

Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Caraterização de vírus específicos de insetos: análise da sua diversidade genética e relação com outros vírus

Autor: Paulo Jorge Sampaio Morais

Orientador: Prof. Dr. Ricardo Parreira

Coorientador: Profª. Drª. Ana Abecasis

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de

Doutor em Ciências Biomédicas, especialidade em Microbiologia

Oral/Poster communications:

- **Morais, P.** & Parreira, R. (2019) "Primeira descrição de flavivírus e densovírus específicos de insetos em mosquitos de Angola: deteção de genoma e caracterização filogenética de sequências virais" December 12. *X Jornadas do Instituto de Higiene e Medicina Tropical*. Poster presentation.
- **Morais, P.** & Parreira, R. (2020) "Dispersão espaciotemporal de flavivírus específicos de insetos explorando o uso da sequência codificante da RNA polimerase" December 14. *XI Jornadas do Instituto de Higiene e Medicina Tropical*. Online.
- **Morais, P.** & Parreira, R. (2021) "Reavaliação da diversidade genética e taxonomia da família Mesoniviridae, bem como das suas relações com outros nidovírus" December 10. *XII Jornadas do Instituto de Higiene e Medicina Tropical*. Poster presentation.

Peer-reviewed published/accepted manuscripts directly related with the work presented in this thesis:

- Manuel Silva, **Paulo Morais**, Carla Maia, Carolina Bruno de Sousa, António Paulo Gouveia de Almeida, Ricardo Parreira. *A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal.* Virus Research, Volume 274 (July 2019), Article nº 197769. https://doi.org/10.1016/j.virusres.2019.197769

- **Paulo Morais**, João Pinto, Cani Pedro Jorge, Arlete Dina Troco, Filomeno Fortes, Carla Alexandra Sousa, Ricardo Parreira. *Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: Genome detection and phylogenetic characterization of viral sequences.* Infection, Genetics and Evolution, Volume 80 (January 2020), Article nº 104191. https://doi.org/10.1016/j.meegid.2020.104191

- **Paulo Morais**, Nídia Trovão, Ana Abecassis, Ricardo Parreira. *Genetic lineage characterization and spatiotemporal dynamics of classical insect-specific*

flaviviruses: outcomes and limitations. Virus Research, Volume 303 (October 2020), Article nº 198507. https://doi.org/10.1016/j.virusres.2021.198507

- **Paulo Morais**, Nídia Trovão, Ana Abecasis, Ricardo Parreira. *Readdressing the genetic diversity and taxonomy of the Mesoniviridae family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences.* Virus Research, Volume 313 (May 2022), Article nº 198727. https://doi.org/10.1016/j.virusres.2022.198727

- **Paulo Morais**, Nídia Trovão, Ana Abecasis, Ricardo Parreira. *Insect-specific viruses in the Parvoviridae family: Genetic lineage characterization and spatiotemporal dynamics of the recently established Brevihamaparvovirus genus.* Virus Research, Volume 313 (May 2022), Article nº 198728. https://doi.org/10.1016/j.virusres.2022.198728

In honor of my father, grandfather and grandmother, who all unfortunately passed away during this project…

"Think of him faring on, as dear

In the love of There as the love of Here.

Think of him still as the same. I say,

He is not dead—he is just away."

- *James Whitcomb Riley*

Acknowledgements

- To IHMT for allowing me to continue my studies and further pursue my lifelong wish of obtaining a doctorate degree, and to GHTM for providing the bioinformatics framework that supported the majority of the analyses here presented.
- To my supervisor, Ricardo Parreira, for the constant guidance, patience and support, for always being available and for all the handed experience and knowledge that will for sure be invaluable in the future. Faced with all my mistakes and mishaps, he never once doubted my work and sought for all projects/experiments to provide the best possible results.
- To my co-supervisor, Ana Abecasis, who, even through her busy schedule, always provided the best support she could.
- To the members of my Tutorial Commission, Nídia Trovão and João Piedade, for all assistance over the project.
- To the coordination of my doctoral program, formerly João Pinto and now Gabriela Santos-Gomes, for allowing me to pursue this dream and finish this project.
- To all my colleagues at IHMT, for all the friendship.
- To my wife Sara, who goes through all my joy and, more importantly, my suffering. She's the rock that keeps me going forward and without her I would probably not even have embraced this challenge. Your love and support were essential through all these years.
- To my father, Vitor Morais, who passed away on the 29th of January, 2019. Our relationship was not always perfect, but if I am the man I am today, close to finishing this doctoral degree, it's no doubt also thanks to his support. I'm sure he's looking proud at what I have achieved, and I'll keep striving and achieving higher heights for him.
- To my grandfather, Agostinho Morais, who passed away on the 3rd of November, 2018, and my grandmother, Maria de Jesus Morais, who passed away on the 26th of November, 2021. Their emotional, financial and heartfelt support was one of, it not the most, important over the course of not only this doctoral program, but also all my former academic courses. Both were already quite sick when I started this project, but they still wanted to see me cross the finish line. Unfortunately, that did not come to be. I hope I made you both proud.

- To my family, who went through both good and bad moments with me over the last years, and also my wife Sara’s family, who always took me literally as one of their own and always gave all their hearts for me.

Abstract

Arboviruses are responsible for impactful viral diseases and are transmitted to humans and other animals by insect and other arthropod vectors. Among the latter, mosquitoes pose as one of the most important hematophagous vectors known to carry pathogenic arboviruses for humans, like dengue and Zika. For a long while, virus research was driven by the negative impact viruses impose on human/animal/plant health, but recent developments in next-generation sequencing and bioinformatics have allowed for improvements on virus discovery strategies. In turn, these have led to an increase in the number of peculiar viruses detected in viral screening-based studies, some of which display restricted replication in vertebrate cells. These viruses, designated insect-specific viruses, are thought to have none-to-low impact on animal health and have endured in the shadow of pathogenic arboviruses. However, in the last decades they have become the focus of our attention, not only due to their extensive diversity and unusual host-restriction strategies, but also because of their potential to interfere with the replication of arboviruses. Insect-specific viruses have since been discovered in multiple virus families, with mosquito-specific viruses especially associated with the *Flaviviridae*, *Mesoniviridae* and *Parvoviridae* families.

In this project we sought to detect and analyze new insect-specific virus sequences from these three viral families in mosquitoes collected in Portugal, Angola and Mozambique. Genetic diversity, phylogenetic reconstruction and phylodynamic assessments were then executed, using both new generated sequences and sequences available in public databases. New classical insect-specific flavivirus (cISF) sequences were detected in mosquito pools from these three geographic regions, and different sub-lineages inside the cISF cluster were characterized. Phylodynamics analyses suggested that cISF dispersion over space and time could be recent and quite dynamic. On the other hand, while insufficient data did not allow for a full phylodynamic analysis based on mesonivirus sequences, an extensive taxonomy revision was performed, that also included the analysis of sequences similar to mesoniviruses (*meson*-like viruses) recently detected in organisms other than mosquitoes. Finally, we also sought to analyze among the parvoviruses of invertebrates those included in the *Brevihamaparvovirus* genus, that have been restricted (so far) to a few mosquito species. Their genomes seem to evolve under strong purifying selection and are also characterized by low entropy, as also observed for flaviviruses and mesoniviruses. We also performed a taxonomic revision of the *taxon* (the first ever for brevihamaparvoviruses), and attempted their first ever phylodynamic reconstruction.

Keyword: Insect-specific virus; taxonomy; genetic diversity; phylogenetic analysis; spatiotemporal dynamics

Resumo

Os arbovírus são responsáveis por doenças com impacto significativo na saúde humana, sendo transmitidos a humanos e outros animais por insetos e outros vetores invertebrados. Entre estes últimos, os mosquitos representam um dos mais importantes vetores conhecidos por servirem de vetores a arbovírus patogénicos para os humanos, de que são exemplos os vírus da dengue e Zika. Por muito tempo, a pesquisa de vírus foi impulsionada pelo impacto que estes agentes impõem à saúde humana/animal/plantas, mas os desenvolvimentos nas últimas décadas nos domínios das tecnologias de sequenciação de alto rendimento e bioinformática permitiram melhorias nas estratégias de descoberta de vírus, o que, por sua vez, levou a um aumento no número de vírus peculiares que vieram a ser descobertos em rastreios virológicos, alguns com replicação restrita em células de vertebrados. Acredita-se que esses vírus, designados vírus específicos de insetos, tenham impacto nulo ou baixo na saúde animal e, provavelmente por isso mesmo, tenham permanecido à sombra de arbovírus patogénicos. No entanto, nas últimas décadas eles tornaram-se no foco da nossa atenção, não apenas pela sua extensa diversidade e estratégias incomuns de replicação restrita nalguns hospedeiros, mas também pelo seu potencial de interferir na replicação de arbovírus. Desde então, diversos vírus específicos de insetos foram descobertos em várias famílias de vírus, com vírus específicos de mosquitos associados especialmente às famílias *Flaviviridae*, *Mesoniviridae* e *Parvoviridae*.

Neste projeto procurou-se detetar e analisar novas sequências de vírus específicos de insetos dessas três famílias virais em mosquitos coletados em Portugal, Angola e Moçambique. A diversidade genética, reconstrução filogenética e avaliações filodinâmicas foram então executadas, usando tanto sequências genómicas geradas de novo, bem como de sequências disponíveis em bases de dados públicas. Novas sequências de flavivírus específicos de insetos ditos "clássicos" (cISF) foram detetadas em *pools* de mosquitos das três regiões geográficas, e diferentes sub-linhagens de cISF foram caracterizadas. A sua análise filodinâmica sugeriu que a dispersão de cISF no espaço e no tempo deverá ser recente e bastante dinâmica. Por outro lado, embora dados insuficientes não tenham permitido uma análise filodinâmica completa com base em sequências de mesonivírus, foi realizada uma extensa revisão taxonómica, que incluiu a análise de sequências semelhantes a mesonivírus (*meson-like viruses)* recentemente detetadas em outros organismos que não mosquitos. Por fim, também procurámos analisar entre os parvovírus de invertebrados aqueles que têm sido incluídos no género *Brevihamaparvovirus*, cuja distribuição até ao momento parece ser restrita a algumas espécies de mosquitos. Os seus genomas parecem evoluir sob forte seleção negativa e também são caracterizados por baixa entropia, tal como foi igualmente observado para flavivírus e mesonivírus. Também realizámos uma revisão taxonómica do táxon (o primeiro para brevihamaparvovírus) e efetuámos a sua primeira reconstrução filodinâmica.

Palavras-chave: Vírus específicos de insetos; diversidade genética; reconstrução filogenética; filo dinâmica

Abbreviations

3Clpro	3C-like protease
ω	Omega
aa	Amino acid
AAP	Assembly-activating protein
Ace	Angiotensin-converting enzyme
AcMSV	Aphis citricidus meson-like virus,
aLRT	Approximate likelihood ratio test
AMV	Alphamesonivirus
ATP	Adenosine 5'-triphosphate
BBaV	Bontang Baru virus
BEAST	Bayesian Evolutionary Analysis Sampling Trees
BHP	Brevihamaparvovirus
BF	Bayes factor
BLAST	Basic Local Alignment Search Tool
BlN	Botrylloides leachii nidovirus
C	Capsid
CASV	Casuarina virus
CAVV	Cavally virus
CDC	Centers for Disease Control and Prevention
cDNA	Complementary deoxyribonucleic acid
CFAV	Cell fusing agent virus
CHIKV	Chikungunya virus
cISF	Classical insect-specific flavivirus
CO2	Carbon dioxide
COI	Cytochrome c oxidase subunit I
CoV	Coronavirus
CPE	Cytopathic effect

CPFV	Culex pipiens flavivirus
CTFV	Culex theileri flavivirus
CxFV	Culex flavivirus
DB	Dipteran Brevihamaparvovirus
DDBJ	DNA Data Bank of Japan
DENV	Dengue virus
dhISF	Dual host-affiliated insect-specific flavivirus
DkNV	Dak Nong virus
DKV	Dianke virus
dN	Non-synonymous substitution
DNA	Deoxyribonucleic acid
ds	Dataset
dS	Synonymous substitution
dsDNA	Double-stranded deoxyribonucleic acid
dsRNA	Double-stranded ribonucleic acid
E	Envelope
EEEV	Eastern equine encephalitis viruses
ENA	European Nucleotide Archive
ER	Endoplasmic reticulum
ESS	Effective sample size
EVE	Endogenous viral elements
ExoN	Exoribonuclease
FCT	Fundação para a Ciência e a Tecnologia
FEL	Fixed Effects Likelihood
FmN	Fathead minnow nidovirus
GMRF	Gaussian Markov random field
GTR	General time-reversible
HanaV	Hana virus

Hel	Helicase
HPD	Highest Probability Density
ICTV	International Committee on Taxonomy of Viruses
IHMT	Instituto de Higiene e Medicina Tropical
Immeso1	Insect metagenomics mesonivirus 1
ISF	Insect-specific flavivirus
ISV	Insect-specific virus
JC69	Jukes and Cantor, 1969
JEV	Japanese encephalitis virus
JTT	Jones, Taylor and Thornton
K80	Kimura, 1980
KADV	Kadiweu virus
kb	Kilobase
kDa	Kilodalton
KHY85	Hasegawa, Kishino and Yano, 1985
KPhV	Kamphaeng Phet
KRV	Kamiti River virus
KSaV	Karang Sari virus
LtM	Leveillula taurica associated alphamesonivirus 1
MAFFT	Multiple Alignment using Fast Fourier Transform
MAYV	Mayaro virus
MBV	Mosquito-borne virus
MCCT	Maximum-Clade Credibility Tree
MCMC	Markov Chain Monte Carlo
MDA5	Melanoma differentiation-associated protein 5
MEGA	Molecular Evolutionary Genetics Analysis
MenoV	Meno virus
ML	Maximum-likelihood

MM	Mosquito mesoniviruses
MRCA	Most Recent Common Ancestor
mRNA	Messenger ribonucleic acid
MSA	Multiple sequence alignment
MVEV	Murray Valley encephalitis virus
NDiV	Nam Dinh virus
NgeV	Ngewontan virus
NGS	Next-Generation Sequencing
NHUV	Nhumirim virus
NIEV	Niénokoué virus
NJ	Neighbor-joining
NKV	No-known vector virus
nm	Nanometer
NMT	N7-methyltransferase
NP	Nucleoprotein
NS	Non-structural
NseV	Nse virus
nt	Nucleotide
OdoV	Odorna virus
OFAV	Ofaie virus
OM	Other mesoniviruses
OMT	Nucleoside-2'-O-methyltransferase
ORF	Open reading frame
OROV	Oropouche virus
PARV	Psammotettix alienus reovirus
PCR	Polymerase chain reaction
PED	Pairwise evolutionary distance
pI	Osoeletric point

PM	Parsimony method
PP	Posterior probability
prM	Pre-membrane
pRT	Para-retroviruses
PS	Path Sampling
RDP	Recombination detection program
RdRp	RNA-dependent RNA polymerase
RIG-1	Retinoic acid-inducible gene I
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RRW	Relaxed random walk
RT	Retroviruses
RT-PCR	Reverse transcription polymerase chain reaction
RVFV	Rift Valley fever virus
S	Spike
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SAP	Single amino-acid polymorphism
SAT	Small alternatively translated protein
SheV	Simian hemorrhagic encephalitis virus
SLAC	Single Likelihood Ancestor Counting
SMART	Simple modular architecture research tool
SNAP	Screening for non-acceptable polymorphisms
SS	Stepping-Stone
ssDNA	Single-stranded deoxyribonucleic acid
ssRNA	Single-stranded ribonucleic acid
TBE	Tick-born encephalitis
TBV	Tick-borne virus
TLR3	Toll-like receptor 3

TMUV	Tembusu virus
TN93	Tamura-Nei, 93
USA	United States of America
UTR	Untranslated region
VIGS	Virus-induced gene silencing
VP	Viral protein
WAG	Whelan And Goldman
WEEV	Western equine encephalitis viruses
WHO	World Health Organization
WNV	West Nile virus
YFV	Yellow fever virus
ZIKV	Zika virus

Table of Contents

List of figures

List of tables

Chapter 1. General Introduction

1. Viruses

Viruses are submicroscopic, obligate intracellular parasites, composed of genetic information (in the form of RNA or DNA), comprising genes encoding proteins dedicated to self-replication (sometimes also including those that encode products that change the cellular environment to benefit viral replication and/or to deter host antiviral defenses), as well as those that encode the structural components that determine the formation of the infectious virus particle itself (the virion). In addition to the latter, sometimes a lipid-rich envelope (derived from one of the multiple cellular membrane compartments), encloses the viral genome inside a viral particle (Wagner et al., 2007a). On their outer surface, regardless of the presence/absence of an envelope, the viral particle displays the proteins that are essential to allow the virus to interact with the host cell (Mothes et al., 2010).

While viruses are ubiquitous and structurally simple (Cann, 2015), they can come in many shapes and sizes (as seen in Figure 1), which vary between different viral families. Viruses can be classified according to their type of genome and replication strategy (Baltimore Classification System) and can also be recognized in different taxonomic ranks (like different viral families), which focuses mainly on the grouping of closely related viruses. Official taxonomic classification has been overseed by The International Committee on Taxonomy of Viruses (ICTV) and has evolved over time. It started as a five-rank hierarchy (from species to genus, subfamily, family and order) and a simple phenotype-based characterization, to a 15-rank structure that also includes the comparative analysis of sequences of conserved genes and proteins (Gorbalenya et al., 2020). Changes to the taxonomic classification of viruses, following changes to the ICTV Code in 2018, can be provided by all of the virology community (Siddell et al., 2019).

Until a few decades ago, the isolation and study of such agents were mostly focused on those that were pathogenic to humans, or to the animals/plants humans depend upon for their survival or with significant associated economic value. Unsurprisingly, special attention was devoted to viruses that led to epidemic/pandemic events, where the analysis the investigation of their characteristics, and not only their pathogenicity/virulence, promoted the evolution of virology (Oldstone, 2019).

Virology encompasses, amongst other aspects, the study of viral structure and classification, their interactions with other organisms (viruses use as hosts, natural

reservoirs, or vectors to warrant their transmission), their evolution as well as genetic diversity, ecology, and the mechanisms they explore to ensure their successful replication inside a host cell (Wagner et al., 2007b). Still, the demanding processes involved in the discovery and isolation of viruses, as well as a focus on disease-associated viral agents, proved as a limitation in the study of the diversity of the virosphere, from which only a very small fraction is still currently known (Call et al., 2021). However, in recent years, research in virology has mainly been approached from a molecular biology perspective, and many studies have been especially focused on the analysis of viral genomes using a combination of metagenomic approaches, taking advantage of the augmented performance in high throughput sequencing and exploring the developments in the field of bioinformatics. In combination, these have been vital in improving our current knowledge about the viral diversity of the biosphere (Ibrahim et al., 2018). This has proven especially true when our attention is focused either on rarely sampled taxa or infrequently visited biotopes, as viral surveys have repeatedly revealed novel or divergent viral groups (Li et al., 2015; Kauffman et al., 2018). Invertebrates are among the animals most frequently sampled in recent viral surveys, and among them, mosquitoes (Diptera: *Culicidae*), due to their role as vectors of pathogenic viruses, are the invertebrates most frequently studied (Zhang et al., 2018b).

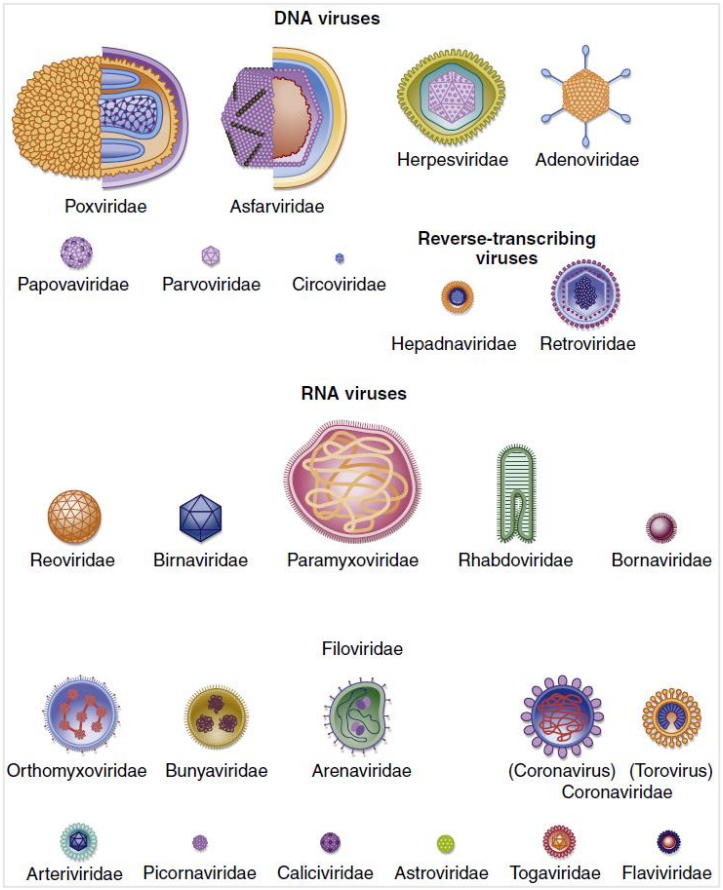


Fig. 1: Illustration of examples of different viruses' shapes and sizes, based on the family they belong to (Retrieved from Cann, 2015). Retrieved from Principles of Molecular Virology (6th Edition) by Alan J. Cann, published by Elsevier, Copyright © Academic Press, 2015. Permission to reusage granted by publisher.

1.1. Arthropod-borne diseases and arboviruses

Many viral diseases are caused by viruses carried by insect vectors, with human involvement often being incidental (i.e., humans do not always participate in the natural maintenance cycles of the viruses in question). These viruses, identified as arthropod-borne viruses (or arboviruses) are usually biologically maintained in a natural cycle that includes a vertebrate and a virus-infected invertebrate vector. These viruses have imposed great challenges on humans, and for many of them, their disease-causing mechanisms are still unknown. Moreover, they have proven resilient to the different strategies implemented toward disease control, from difficulties regarding the limitation of the geographical dispersal of their vectors, to the development of antivirals and vaccines (Young, 2018).

Vector-borne diseases are infections that list among the preeminent causes of morbidity and mortality in humans, causing more than 700,000 deaths annually (WHO, 2020), and constitute a significant health problem, especially in tropical and subtropical countries (Figure 2). Recent years saw a considerable expansion of arboviral diseases linked to multiple factors that include high human population density in certain areas, trade globalization, and climate changes (Martina et al., 2017). Even after the onset of COVID-19, their impact on human health has been largely disregarded. Viruses involved in these diseases are either primarily, or exclusively, transmitted through a hematophagous invertebrate (Cann, 2015), with the most common vector-borne diseases being associated with either insects (mostly mosquitoes) or arachnids (mostly ticks). Arbovirus transmission mechanics are quite complex and can be sorted into three main types, which mainly involve either humans or other vertebrates (Weaver et al., 2018):

- **Direct spillover**: direct virus transmission to humans by primary/surrogate enzootic vectors usually associated with enzootic amplification, augmenting viral circulation near humans. Examples include the West Nile virus (WNV), St. Louis encephalitis virus, yellow fever virus (YFV), as well as the Eastern and Western equine encephalitis viruses (EEEV and WEEV, respectively);
- **Domestic animals as amplifiers**: prior amplification in domestic animals, succeeded by direct spillover to humans. Examples include the Japanese encephalitis virus (JEV; with viral amplification in swine) and Rift Valley fever virus (RVFV; with viral amplification in sheep);
- **Enzootic transmission cycle to a human–mosquito–human cycle**: an enzootic cycle that spills over to infect people that live nearby, who then serve as amplification hosts. Examples include the Zika virus (ZIKV) and chikungunya virus (CHIKV).

Mosquitoes are insect vectors that are known to transmit pathogenic agents with emerging potential, including viruses (Gould et al., 2017). Multiple mosquito-borne viruses, like WNV, CHIKV, dengue virus (DENV), YFV, and ZIKV, pose a substantial potential threat to public health (Huntington et al., 2016). Among the latter, DENV stands as the most important human arbovirus. Genetic studies have suggested that its origin may have occurred around a thousand years ago, but it was isolated for the first time in Japan only

in 1948. Presently, it is associated with significant morbidity and mortality, especially in low-income countries (Harapan et al., 2020). Other than the most well-known (or talked about) arboviruses, multiple others have remained in the shadows of the most predominant ones and have either been emerging at an alarming rate or are poised to do so should conditions allow it. One such example is the Tembusu virus (TMUV), belonging to the *Flaviviridae* family. While TMUV was first detected in Malaysia in 1955, it was not until 2010, when the first major outbreak occurred in ducks, that research focused on this virus really bloomed (Hamel et al., 2021). Recent years saw the emergence of new arboviruses, like the Chatanga virus, first detected in Finland in 2007 (Madani et al., 2011), and the Kibale virus, first detected in Uganda in 2013 (Marklewitz et al., 2013), both mosquito-borne viruses from the family *Peribunyaviridae*. Recent years also saw many well-known arboviruses reemerge (including CHIKV and ZIKV), even leading to their introduction from endemic areas into new regions, including Europe (Barzon, 2018).

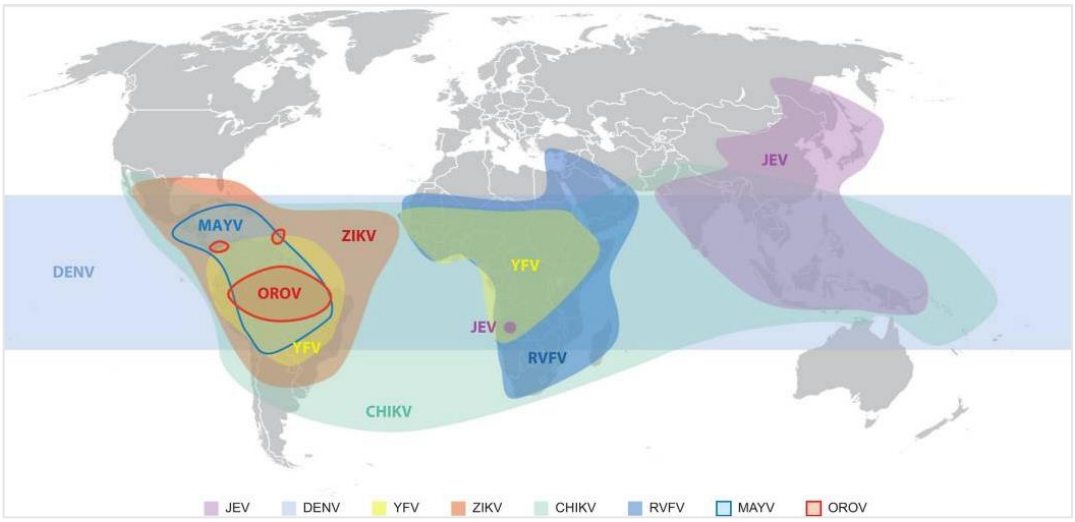


Fig. 2: Reported distributions of arboviruses, as Weaver et al. (2018) reported. Abbreviations: CHIKV - chikungunya virus; DENV - dengue virus; JEV - Japanese encephalitis virus; MAYV - Mayaro virus; OROV - Oropouche virus; RVFV - Rift Valley fever virus; YFV - yellow fever virus; ZIKV - Zika virus. Reprinted by permission from Microbiology Society under the license number 1202419-1, from 27 Apr 2022.

There is a lot to be learned from the analysis of insect-borne viruses. Even with the recent resurgence of multiple arboviruses, only less than 1% of all viruses, by recent estimates, have been discovered so far (Geoghegan & Holmes, 2017). Unsurprisingly, starting

approximately a decade ago, some viral screening-based studies combining metagenomics and Next Generation Sequencing (or NGS) started reporting newly discovered insect-associated viruses, many of which did not correspond to *bona fide* arboviruses, as their replication was restricted to invertebrates or invertebrate cell-lines (Junglen & Drosten, 2013).

1.2. Insect-Specific Viruses

In recent years, a large assemblage of viruses that seem to be non-pathogenic to vertebrates, and that are unable to replicate in vertebrate cells, have come to be generally known as insect-specific viruses (ISVs). They are abundant in hematophagous arthropods, especially in mosquitoes, and comprehend a genetically disparate assembly of RNA and DNA viral agents, belonging to several different families (Nouri et al., 2018).

1.2.1. History of ISVs

The first-ever ISV to be described was isolated from an *Aedes aegypti* cell culture in 1975 by Stollar and Thomas (Stollar and Thomas, 1975). No cytopathic effects were observed when the virus was inoculated in different vertebrate cell lines, and it would eventually be characterized as a positive-sense RNA virus, and placed within the family *Flaviviridae*. Furthermore, it was designated cell fusing agent virus (CFAV) after the typical cytopathic effect it causes when it replicates in mosquito cell-lines, which involves the formation of multiple large syncytia (Cammisa-Parks et al., 1992). More than 20 years passed with little to no developments regarding the discovery of new ISVs, and it was not until the early 2000s that ISV research was solidly boosted, starting with the description of another insect-specific flavivirus, the Kamiti river virus (KRV; Crabtree et al., 2003).

Interest in the analysis of ISVs mounted at a slow pace, with research showing these genetically diverse viruses belonged to different *taxa*. In recent years, a clear rise in their study has been indisputable (Abudurexiti et al., 2019), benefiting from the widespread use of molecular technologies and the development of bioinformatics tools, which have granted us a better understanding of the nature, diversity, distribution, replication features and evolution of ISVs (Nouri et al., 2018). Indeed, studies have already shown that at

least for some ISV, their genetic information not only exist in the form of viral genomes but could also be found integrated into the genomes of mosquitoes where it might have remained for a long time, and, therefore contribute to the acquisition of genetic diversity in eukaryotic cells. Specifically, initial studies reported partial sequences related to the ones encoding the RNA dependent RNA polymerase of CFAV and KRV in the *Aedes aegypti* A20 cell-line (Crochu et al., 2004).

Even after the sudden surge of ISV discovery in the early 2000s, it was not until early 2010s that ISV research really spiked. Innumerous viruses would be identified, the majority with RNA genomes. While they frequently corresponded to new flaviviruses (Order *Amarillovirales*), others were placed in multiple families in the Order *Bunyavirales* (Hobson-peters et al., 2016), the family *Mesoniviridae* in the Order *Nidovirales* (Wang et al., 2017), and also in the families *Togaviridae* in the Order *Martellivirales* (Nasar et al., 2012) and *Rhabdoviridae* in the Order *Mononegavirales* (Ma et al., 2014). However, the taxonomic diversity involving ISVs would grow significantly in recent years. Indeed, da Silva et al., in 2020, detected ISVs from innumerous different virus families from mosquitoes collected in Brazil, including *Circoviridae, Totiviridae, Iflaviridae, Nodaviridae, Luteoviridae, Phasmaviridae, Phenuiviridae, Rhabdoviridae, Orthomyxoviridae* and *Xinmoviridae*. Also in 2020, a double-stranded RNA (dsRNA) virus, denominated *Psammotettix alienus* reovirus (PARV), was isolated from a leafhopper (*Psammotettix alienus*) collected in China (Fu et al., 2020). While ISVs do not usually display a DNA genome, a large number of ISV sequences have also been associated with members of the *Parvoviridae* family (Zhai et al., 2008). Recent research does suggest ISVs have a broader distribution that initially anticipated, not only being associated with diverse virus families, but also displaying a global distribution. Indeed, ISVs from the *Flaviviridae*, *Mesoniviridae* and *Parvoviridae* families have been discovered in multiple continents (Sadeghi et al., 2017; Shi et al., 2017; Kyaw Kyaw et al., 2018), and in association with distinct Culicine and Anopheline mosquitoes (Calzolari et al., 2016).

Over time, ISVs have also been described in the research work carried out at IHMT, which allowed for the characterization of such viruses from mosquitoes collected in Portugal. As examples, the first world-wide known insect-specific flavivirus (ISF) from *Culex theileri* was identified in mosquitoes collected in southern Portugal (Parreira et al.,

2012). This was followed by the full characterization of an ISF isolated from *Ochlerotatus caspius* (Ferreira et al., 2013), as well as Negev-like viruses (Carapeta et al., 2015), also found in mosquito specimens from the south of Portugal. Further research in 2019 (based, once more, on mosquitoes collected in 2018 in the Algarve region) allowed for the detection of ISVs from the *Flaviviridae*, *Parvoviridae* as well as unclassified viral sequences, associated with putative viruses placed in the Order *Bunyavirales* (Silva et al., 2019).

1.2.2. Transmission and host-range restrictions of ISV

How insects acquire ISVs, and how these viruses are maintained in nature, is still largely unknown. Unlike arboviruses, which are usually acquired by a vector upon feeding on a viremic vertebrate, there is no indication that ISVs can be natural maintained in a cycle between a mosquito and a vertebrate animal (Weaver & Barrett, 2004). While different types of viral transmission modes could be explored to preserve ISVs in their host populations, the primary natural maintenance mechanism is thought to be vertical (or transovarial) transmission. This is suggested by direct experimental evidence, where the virus was shown to pass from infected female mosquitoes to their offspring (Haddow et al., 2013; Saiyasombat et al., 2011). Venereal transmission (transmission from naturally infected male mosquitoes to females) has also been suggested and could be a possible mechanism of ISV maintenance (Bolling et al., 2012).

While the story of the discovery of ISVs is relatively easy to track, the mechanisms behind host restriction appear to be way more complex. Several experimental studies in recent years confirmed ISV host-restriction by attempting their propagation, for example, in vertebrate cell lines, including unsuccessful attempts to replicate *Culex* insect-specific flaviviruses in African green monkey kidney epithelial cells (Hoshino et al., 2007) and chicken embryo fibroblasts (Bolling et al., 2011). Also, studies have shown that direct transmission in nature of these viruses between infectious insects with ISV and vertebrates could be limited, as reports have shown that some ISVs, like a *Culex* flavivirus (CxFV) and the Palm Creek virus, could not be found in the saliva of *Culex* mosquitoes (Hall-Mendelin et al., 2016; Talavera et al., 2018). Surprisingly, one study did report salivary glands of *Aedes aegypti* mosquitoes infected with Eilat virus, an insect-specific

alphavirus (Nasar et al., 2014), which means ISVs could eventually come into contact with vertebrate cells.

How could we then explain the limited *in-vitro* replication of ISVs in vertebrate cells lines that have been reported in multiple experimental data? Several studies highlighted that host-range restrictions of ISVs in vertebrate cells may affect multiple stages of the viral replication cycle (Junglen et al., 2017). This was demonstrated when a YFV chimera carrying the envelope proteins of an ISV, the Niénokoué virus (NIEV), exhibited efficient replication in invertebrate cells, though it could not enter vertebrate cells. When RNA from the YFV/NIEV clone was inserted into those same vertebrate cells, no viral genome replication nor assembly was recorded. One of the most supported theories suggest that the innate immune system could strongly hinder ISV replication in vertebrate cells. Tree et al., in 2016, suggested that the ISV Kamiti River virus (KRV) could not evade vertebrate innate immune pathways. Knockdown of pattern recognition receptors (RIG-1, MDA5 and TLR3) resulted in a rise of KRV replication, even if at low levels, in vertebrate cells, which suggests KRV can replicate in those same cells if the innate immunity pathways are silenced. Other reports observed that even micro ambient temperature could be an important factor determining host restriction. While arboviruses are capable of replicating at temperatures up to 42 ºC, insect-specific viruses seem to only be able to replicate at ambient temperatures (between 25-28 °C). On the other hand, Marklewitz et al. (2015) observed that high temperatures hindered the replication of an insect-specific bunyavirus in insect cells, and also that simply lowering the temperature was not enough for that same ISV to replicate in vertebrate cells.

While there is no definitive answer to the host-restriction of ISVs, these data suggest that host restriction could be explained by either ISVs genetic elements, vertebrate host factors or even be dependent on micro ambient temperature.

1.2.3. ISVs potential for vector control

The global threat that arboviruses present, the lack of efficient treatment protocols as well as prophylactic vaccines against many of them, has led, in recent years, to research on new mechanisms for disease control. Since very little has been achieved regarding the

development of antivirals, or even vaccines for the majority of arboviruses, the control of arboviruses has mainly targeted mosquito populations (Ache et al., 2019).

It has been shown that *Wolbachia*, a common intracellular bacterial endosymbiont, can block the transmission of important arboviruses (Jiggins, 2017). While recent research results look promising (Aliota et al., 2016; Indriani et al., 2020), mechanisms by which *Wolbachia* prevents arbovirus transmission are not yet still fully understood. Virus block could be caused by decreasing host virus transmission (possibly by higher *Wolbachia* density), especially in the midgut and salivary glands. In turn, this could lead to arboviruses replication inhibition (Martinez et al., 2014). Alternatively, *Wolbachia* could decrease host population density if its presence somehow has negative fitness effects in mosquitoes. The latter effect has been investigated in mosquitoes with experimentally-introduced *Wolbachia*, but mixed results have been obtained so far (Ross et al., 2019).

In theory, regardless of the impact ISVs might have on the fitness of mosquitoes (see below), they could potentially be used as biocontrol agents by exploring their symbiotic relationship with their mosquito hosts, which, in turn, could eventually lead to inhibition of arbovirus replication in ISV-infected cells (Calisher and Higgs, 2018). Indeed, experimental studies already demonstrated ISVs could modulate pathogenic arboviruses replication in mosquitoes (Fujita et al., 2018). The main mechanism that has been proposed for a plausible application of ISVs in arbovirus control strategy is thought to involve superinfection exclusion (Laureti et al., 2020). This implies that a primary infection by an ISV could result in the modification of cell surface molecules that might restraint cell entry by an exogenous virus, or even affect viral intracellular replication, therefore hindering vector competence. The latter is quite complex and usually refers to the ability of the host to withstand infection and subsequently maintain and transmit an infectious agent, usually further involving interactions between host, vector and pathogen and calculated by both insect species factors (longevity, feeding habits) and environmental factors.

Goenaga et al. (2015) demonstrated that a concurrent infection of *Aedes albopictus* C6/36 mosquito cells with the Nhumirim virus (NHUV), an insect-specific flavivirus, and WNV, resulted in a substantial reduction in the replication of the latter. This does suggest that the NHUV reduced vector competence for WNV in the mosquito host. Other studies

also demonstrated that a persistent *Culex* insect-specific flavivirus infection in *Culex pipiens* could hinder early replication of WNV (Bolling et al., 2012), or that Palm Creek virus infection of *Coquillettidia xanthogaster* could hinder replication of WNV and Murray Valley encephalitis virus (MVEV) (Hobson-Peters et al., 2013). However, these observations may depend on specific viral combinations, as another study did not show any promising effects in inhibiting the growth of ZIKV with the insect-specific Palm Creek virus in a co-infection of *Aedes* mosquitoes (Koh et al., 2021). In a similar way, a study from 2018 demonstrated that NHUV infection of C6/36 cells could inhibit ZIKV and DENV-2 replication, but not CHIKV (Romo et al., 2018). Indeed, no consistency has been found in recent research, implying that the mechanisms behind a possible effect of co-infection on vector competence could be more complex than initially thought.

Previous infection of insects with ISV could affect vector competence and catalyze desirable effects to impact the transmission of pathogens in mosquitoes in multiple ways (Öhlund et al., 2019; Figure 3). One possibility is that ISVs could directly cramp arbovirus cell attachment by blocking/downmodulating the expression of specific receptors (Figure 3-1). This interaction has been reported multiple times, including with *Anopheles*-specific ISVs (Colmant et al., 2017), but how it was brought about is largely unknown. Transfecting cultured cells with recombinant ISVs designed to express the ligands which could bind to and saturate such receptors could be a viable approach. Another possibility could involve the use of recombinant ISVs containing effector sequences derived from the target pathogenic virus, which would promote resistance to homologous pathogenic viruses, affecting vector competence (Figure 3-3). This strategy has been applied successfully using a recombinant Sindbis virus (an alphavirus found in both insects and vertebrates) to initiate resistance towards DENV in C6/36 cells (Adelman et al., 2001), but has not yet been explored in practical terms with ISVs.

Moreover, it is also unknown to what extent ISVs may affect the fitness of their invertebrate hosts, and potentially contribute, for example, to reduce the longevity of adult mosquitoes. Using ISVs as a means to introduce RNA-induced silencing complexes (RISC) into the host genome, negatively impacting insect survivability and, in turn, affecting vector competence to arboviruses, could be a viable solution (Figure 3-2). Recombinant ISVs could be used to express insect RNA sequences that would target host-specific genes and affect, for example, their development or fecundity. This strategy has

been used successfully for plant defense against pathogenic viruses but it has not yet been explored with arthropods (Bachan & Dinesh-Kumar, 2012).

These proposed approaches could represent a turning point in vector control strategies and consequent modulation of the replication of pathogenic arboviruses in insect cells. However, since ISV research is relatively recent, many questions remain before it may be possible to use any of the above-mentioned strategies (or others) to efficiently target insect vector viability and/or the replication of viral pathogens. How can we deliver ISV to target hosts? Will recombinant ISVs spread efficiently in their hosts' populations? Will their intended desirable effects endure long enough so as to allow them to become efficient in restraining vector competence and/or longevity? Interestingly, endogenous elements from ISVs have already been found in specific hosts' genomes, where they integrate most probably taking advantage of reverse transcriptase activities encoded by retrotransposons present in mosquito genomes (Crochu et al., 2004; Lequime & Lambrechts, 2017; Abílio et al., 2020). Still, a greater understanding of the basic biology and genetics of ISVs is required, and questions such as (i) what are their species-specificity and geographic range, (ii) how diverse are their genomic sequences on an individual and population level, (iii) how high are their mutation rates, (iv) what kind of selective pressure is acting on their genomic sequences (evolving in a diversifying or purifying way), and (v) are there recombinant ISVs already spreading among insect hosts populations, are still open to discussion, and deserve being investigated.

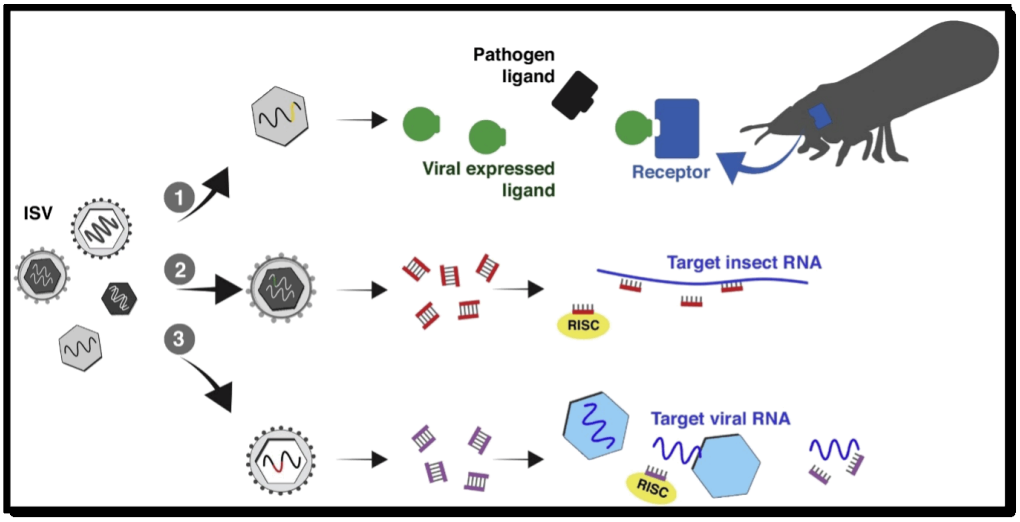


Fig. 3: Potential translational applications of recombinant ISVs by affecting vector competence. Three possibilities for inducing desirable effects in the insect host are indicated: (1) directly impact on pathogen entry in insect hosts by saturating receptors essential for virus acquisition; (2) impact on host survivability by integrating an RNA induced silencing complex (RISC) that target specific host genes and induce adverse effects in insect' essential traits; (3) contribution to increased resistance to the target virus in the insect vector by using recombinant ISVs that contain sequences derived from the target pathogenic virus. Retrieved from Nouri et al. (2019). Reprinted by permission from Elsevier under the license number 5274340001060, from 22 Mar 2022.

In addition to all the questions listed above, one area where ISV research has been lagging regards the analysis of their evolution. How, and for how long, have ISVs been spreading among insect populations? It is known that they are restricted to infect insects, but did they lose the capacity to infect vertebrates over time, or did pathogenic arboviruses arise from ISVs?

1.2.4. Why invest in the study of ISV's viral evolution?

For decades, the application of phylogenetics has been key in allowing the study of the ancestral relationships of viruses and the emergence of viral diversity (Grenfell et al., 2004). Not surprisingly, research on virus evolution has been mainly focused on mosquito-associated viruses with significant impact on public health, including recent epidemic episodes of ZIKV (Pielnaa et al., 2020), YFV (Diagne et al., 2021), WNV (Casimiro-Soriguer et al., 2021) or DENV (Du et al., 2021).

Deconstruction of history, evolution, and relationships among taxonomic operational units, also known as phylogenetic analysis, has been extensively used in multiple areas of biology. As genome sequence data, both in the form of nucleotide and amino acid sequences, became available, research on molecular evolution (which investigates the accumulation of genetic differences over time) has become integral for the analysis of genetic divergence between *taxa*, and to infer the chronology of the splitting events depicted in a phylogenetic tree, or the age of ancestral sequences (Yang et al., 2012). This area of expertise, known as molecular phylogenetics, has been a key element of exploratory and comparative sequence data analysis, which, when applied to virology, allows for the study of relationships among viral genes and the origin and spread of viruses (Yang et al., 2012). The main principle is simple yet also complex.

While it is true that as two sequences diverge from their last common ancestor, so does the number of differences between them increase, simply counting these differences is complicated by multiple factors. These include, for example, the effects of natural selection, the accumulation of differences between sequences sharing common ancestry at variable evolution rates over time or the possibility that multiple substitutions might hit the same nucleotide position (Holder & Lewis, 2003). Also, different viruses, from double-stranded DNA (dsDNA) to single-stranded RNA (ssRNA), display different mutation rates (Figure 4). These result from the combined effects of the biochemical features of the polymerases that replicate their genomes, that may, or may not, introduce substantial numbers of polymerization errors, and the possibility that some of these errors may end-up being corrected during, or after, viral genome replication has been completed, selected for, or wiped out, from the viral population. A clear understanding of the basics of phylogeny is needed to understand how we can use phylogenetic inference and apply it to analyse relationship between ISV genomic data.

Phylogeny allows for the observation of viral molecular evolution through the reconstruction of the so-called phylogenetic trees. These are no more than schematic structures depicting the relationships among the sequences being compared, using a representation where nodes are joined by branches (Figure 5). Each branch represents a relationship between sequences, and each node depicts the birth of a new lineage (or a new individual sequence that will be ancestral to all further sequences in that same node going forward). The sum of this unique ancestor and its respective descendants produces

groups called monophyletic clades. The common ancestor of all sequences in the tree is represented through a root (Yang et al., 2012). Information can then be extracted from branch lengths, where different methods can be applied to calculate the amount of evolution in them, usually expressed in nucleotide substitutions per site or per time unit.

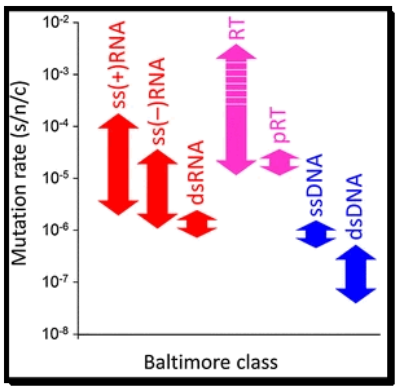


Fig. 4: Mutation rate variation among the seven groups of viruses, according to the Baltimore classification (ss – single-strand, ds – double-strand, RT – retroviruses, pRT – para-retroviruses). Retrieved from Sanjuán & Domingo-Calap (2016). Reprinted from open access article distributed under the terms of the Creative Commons CC BY license.

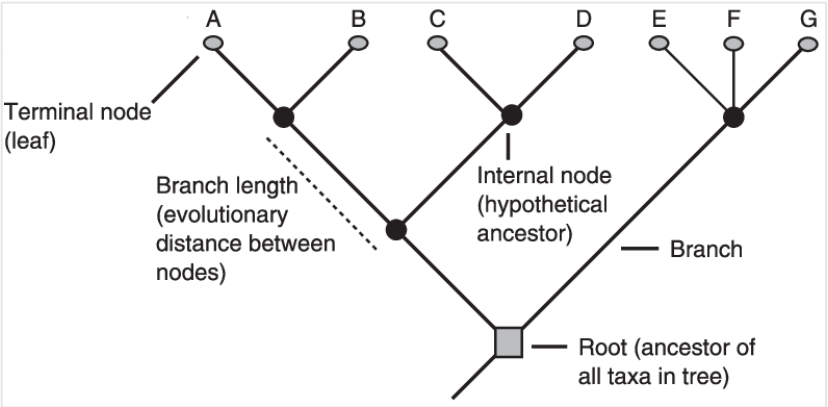


Fig. 5: Example of a phylogeny with included terminology. A and B are considered sister *taxa*, derived from a common ancestral node; all these sequences are inserted into a monophyletic group, including an ancestor with all its descendants. Retrieved from Egan (2006), Copyright © BYU, 2006

Phylogenetic information in datasets can be estimated via likelihood mapping, using the quartet puzzling algorithm (Stirmmer and von Haeseler, 1997), which computes four-taxon trees and applies maximum-likelihood reconstruction to all possible quartets that can be constructed from a specific dataset. A consensus rule is then applied and the output shows the percentage of randomly sampled quartets of sequences that are well resolved (when one of the three possible unrooted tree topologies is favored), along with partially resolved (two tree topologies are equally probable), or not resolved (no single tree topology is favored). The higher the percentage of unresolved quartets, the less befit a dataset is for phylogenetic reconstruction.

Phylogenetic tree reconstruction is a complex process that revolves around, first and foremost, the disclosure/collection and alignment of several different sequences, to ensure the comparison of homologous sequences sites. Then, a reconstruction method must be applied, and they revolve around either so-called "traditional" or Bayesian approaches. Three different reconstruction methods have been repeatedly applied over the years for the reconstruction of phylogenetic trees and in all cases, they attempt to choose the “best” tree possible, each with its own strengths and weaknesses (Holder & Lewis, 2003):

- **Neighbor-joining** (NJ): perhaps, until recently, the most commonly used, and by far one of the quickest, phylogenetic tree reconstruction approaches, is best used for the analysis of sequences with low diversity or that have diverged recently. This method converts nucleotide or protein sequences into a pairwise distance matrix, corrected using one of many possible evolution model formulas. These distances represent the number of changes that occurred along all branches. However, NJ does not hold up well when divergent sequences are being compared, or when homoplasies (resulting from multiple substitutions or back mutations) occur (Frost & Volz, 2013);
- **Maximum-likelihood** (ML): Unlike NJ, ML accurately corrects for multiple mutational events at the same site, analyzing numerous trees resulting from successive topological refinements and choosing the tree that has the highest probability (likelihood) of producing the observed sequences, considering the parameters defined by a chosen evolutionary model (i.e., the ML function maximizes the probability of the data, given a tree and an evolutionary model). It

involves heavy computation and is, therefore, slower than an NJ-based approach. It is, however, recommended to reconstruct relationships between sequences that evolved rapidly or that have split from a common ancestor a long time ago;

- **Maximum Parsimony** (PM): this method also generates scores for each possible tree but, unlike ML, it simply assumes a minimal evolution perspective, and strives at obtaining the tree that reflects the tiniest number of mutations that could possibly produce the observed data. While simple to understand, it does not consider the possibility of different mutational pathways along all branches in the tree.

All the above-mentioned phylogenic tree reconstruction methods require that the chosen tree should be tested for topological reliability using, for example, bootstrapping processes, which are not always easy to interpret (Henderson, 2005). In a departure from the more "traditional" phylogenetic inference approaches, Bayesian methods for phylogenetic reconstruction, instead of searching for the single “best” tree, it considers all sets of plausible, similarly probable trees (weighed by their probability), with the better set being summarized at the end of the analysis as a single Maximum-Clade Credibility Tree (MCCT). One of the greater advantages of Bayesian phylogenetics builds upon the fact that allows for the use of *prior* probability distributions to portray the uncertainty of all unknown parameters before the analysis of the data (including the model parameters). After the data is combined with all possible parameter values and the Markov Chain Monte Carlo (MCMC) is run, the posterior distribution (or probability that the tree is correct, given the data and the chosen evolutionary model) is generated. The popularity of Bayesian methods has risen due to the recent growth of powerful data analysis models and the user-friendly access to computer programs where they have been implemented (Nascimento et al., 2017). Still, since these functions are too convoluted to integrate analytically, Bayesian approaches rely on MCMC algorithms to sample, at random (Markov Chain) from the posterior probability distribution, basing each sample on the previous one (Monte Carlo) (Yang et al., 2012).

As the above sections suggest, choosing the best model for nucleotide substitution calculation is critical in Bayesian phylogenetic approaches, but this is also true for NJ, ML, and MP, to allow analyses to approach biological reality (Yang et al., 2012). The

simplest model to be used is the JC69 (Jukes and Cantor 1969) model, which assumes equal base frequencies and mutation rates. The K80 model, often named Kimura's two parameter model, assumes different rates between transitions and transversions. Both the HKY85 (Hasegawa, Kishino and Yano 1985) and the general time-reversible (GTR) models assume different nucleotide proportions, but while HKY85 allow for one transition rate and one transversion rate, the GTR model assumes a symmetrical substitution matrix with unequal substitution rates for all possible types of substitutions. There are additional models that allow rate variation among different genomic regions of the alignment, usually variable rates according to a gamma distribution (Γ). A model allowing a proportion of sites to be invariable (I) can also be used simultaneously.

Applying phylogenetic approaches to evaluate viral origin and geographic dispersal over time require not only choosing a specific reconstruction approach and the use of an adequate nucleotide substitution model, but also both appropriate coalescent, spatial diffusion, as well as molecular clock models (Figure 6; Pybus & Rambaut, 2009). This field, known as phylodynamics, permits the characterization of the transmission dynamics of virus evolution through the incorporation of epidemiological data alongside molecular sequences (Rife et al., 2017). This analysis can only be possible through the simultaneous application of several analytical models:

- **Molecular clocks**: these models describe the relationship between genetic distance and time (Ho & Duchêne, 2014). The branch lengths only represent genetic diversity between sequences when a simple phylogeny is used, whereas when timestamps are attributed to each known sequence, and evolutionary rates are calibrated using molecular clock models, it becomes possible to estimate the timing of the different branching events along the tree. Unlike older strict molecular clock models, which assumed that all lineages evolved at the same uniform evolutionary rate over time, more recently developed models assume the possibility of using, for example, a local clock (wherein all lineages in a clade share a common substitution rate), or an uncorrelated relaxed clocks (wherein the substitution rate on each lineage is independent from other lineages while being constrained to fit some parametric distribution). As opposed to strict clocks, the latter are generally known as relaxed molecular clocks (Drummond et al., 2006);

- **Coalescent theory**: these statistical models are used alongside molecular clock models (which provide estimates of virus sequence divergence times) to allow simple calculations capable of connecting the demography of a viral population to its sample genealogy, linking patterns of genetic diversity to ecological processes such as population size and growth. Over time, changes in population sizes can be perceived using multiple models, the simplest of which are parametric, where the so-called effective population size may, or may not, vary through time according to a certain function. Nonetheless, it is sometimes desirable to take a more flexible non-parametric approach to demographic modelling (Liu et al., 2009). While the older coalescent models are meaningful only if the sampled population befits the stated demographic model (e.g., during an epidemic when viral population is expected to be expanding exponentially), this is not always the case. As an alternative, flexible nonparametric models, which include, for example, the Bayesian Skyline, Skyride and SkyGrid models, have since been unfolded, enabling estimating varying effective population sizes over time, that don't necessarily fit the shape of previously considered traditional demographic models (Hill & Baele, 2019).
- **Spatial diffusion**: these models consider locality when describing the transmission of a virus, detailing them as agents that move from one place to another. Using sequence sample locations (usually as latitude and longitude coordinates), a Markov chain can be employed to perceive diffusion between locations, allowing the simulation of geographical changes and its integration into phylogenetic and temporal data, allowing spatiotemporal reconstructions. Past spatial diffusion approaches implemented models for discrete transitions in a Bayesian inference framework, allowing geographical information to be blended as distance-informed priors (Lemey et al., 2009). However, these did not explicitly model the diffusion process in continuous space. As such, recent models adopted more relaxed rates by accommodating branch-specific variations i.e., where diffusion rates are drawn independently on each branch of the rooted phylogeny (Lemey et al., 2010).

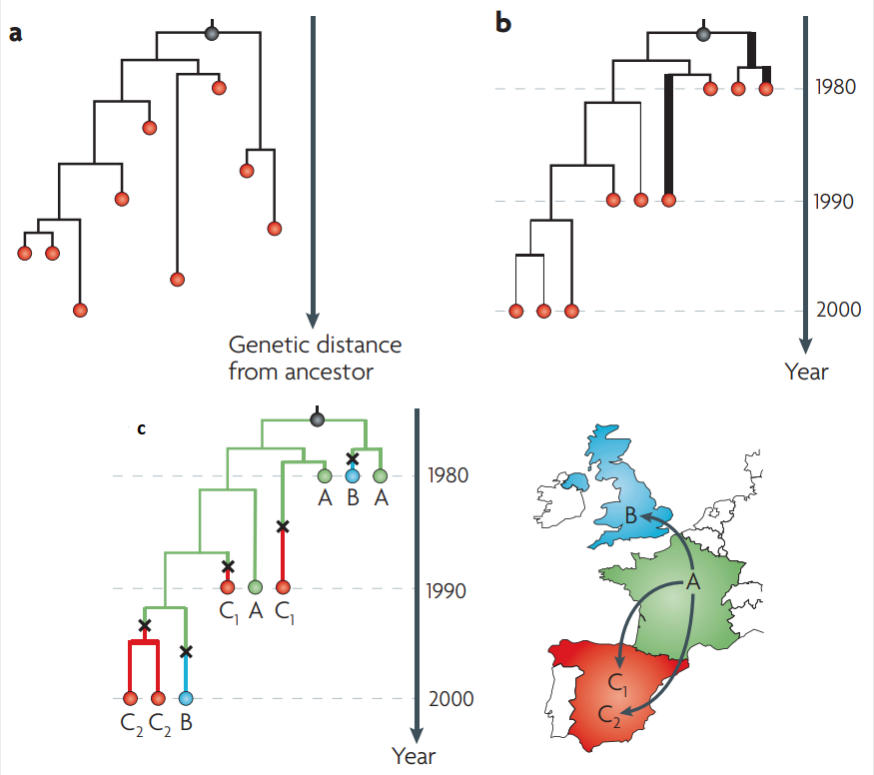


Fig. 6: Phylodynamic processes. (A) Simple rooted phylogenetic tree, with branch lengths representing the genetic divergence from the ancestor (with no timescale); (B) Same tree as A but reconstructed using a molecular clock, which defines a relationship between genetic distance and time, with branch lengths represented in units of years; (C) Same tree as B but reconstructed using spatial data, with each branch labeled as to its estimated geographical position. Combining temporal and spatial data allows further insight into the spatiotemporal dispersal of viruses. This hypothetical virus first spread into France and the United Kingdom, and spatiotemporal data allowed us to identify two different diffusion events into two other locations in Spain, first to C1 in 1990 and later in 2000 to C2. Retrieved from Pybus & Rambaut (2009); Reprinted by permission from Springer Nature under the license number 5232480528990, from 19 Jan 2022.

Before attempting to infer time-scaled trees from a group of sequences, care should be taken in order to confirm if those same sequences possess sufficient temporal signal (genetic changes between sampling times that allow for a reconstruction of the relationship between genetic divergence and time) for a reliable estimation. This can be attained through a regression-based approach through multiple bioinformatic tools, like TempEst, which takes a phylogenetic tree and uses the date for each sequence to analyze root-to-tip genetic distances against sampling time (Rambaut et al., 2016).

1.2.5. Selection of ISV groups

Combining spatial, temporal and genetic analyses could not only reveal the location of origin, and the route of dispersal, of different ISVs, but potentially help predict patterns of future dissemination of these viruses (Pybus & Rambaut, 2009). However, caution is needed when considering which viral families we should analyze, as these researches require heavy computation and heavy data (sequence) generation and/or mining, which is labor-intensive and time-consuming. Still, the nature of the sequences available should also be taken into consideration. While ISVs have already been identified in numerous virus families, some of them are either only known for restricted geographical areas, or are yet represented by a very low number of sequence data in the public genomic databases. In turn, this could hinder sequence analysis due to sampling bias. Indeed, past studies already demonstrated that phylodynamic patterns can be highly impacted by the sampling process (Frost et al., 2015). As a result, research should be focused, whenever possible, on ISV families with the highest number of sequences available, and with a widespread geographical distribution.

With all available information, the core research of this PhD thesis project focused on three specific, and very diverse, groups of ISV *taxa*, which represent specific genera in virus families with the higher representation of mosquito-specific virus sequences in public databases: mesoniviruses (*Mesoniviridae* family, Order *Nidovirales*), brevihamaparvoviruses (*Parvoviridae* family, Order *Piccovirales*) and, finally, insect-specific flavivirus (*Flaviviridae* family, Order *Amarillovirales*), the latter representing the group of ISVs for which the highest known number of genomic sequences are available. Indeed, identification and characterization of ISV sequences from mosquitoes have been especially focused on these three families (Carvalho et al., 2021). While they are all similar in their insect-specific host restriction, they are very distinct when it comes to their genomic features, coding capacity and basic structure (like size and nucleic acid class). In this regard, both flavivirus and mesonivirus possess RNA genomes with sizes of ~11 kb to ~20 kb (respectively), while brevihamaparvovirus have DNA genomes with a smaller size (~4 to ~6 kb).

2. The *Flaviviridae* family

According to the last report by the ICTV, the *Flaviviridae* family is split into four genera: *Flavivirus*, *Hepacivirus*, *Pestivirus*, and *Pegivirus* (Simmonds et al., 2017). The *Hepacivirus* genus is divided into 14 species (from Hepacivirus A to N), including the human Hepatitis C virus, which has since been renamed to Hepacivirus C (Smith et al., 2016). The *Pestivirus* genus currently has 11 species (Smith et al., 2017), including the classical swine fever virus, also known for causing a fatal disease of swine, known as hog cholera (Blome et al., 2017). Pegiviruses belong to a new genus recently proposed in the *Flaviviridae* family, which has been split into 11 species (Smith et al., 2016). Finally, the genus *Flavivirus* is the most distinctive and also most diverse genus of the *Flaviviridae* family, encompassing a genetically distinct array of over 50 RNA viruses, roughly spherical in shape and enveloped, between 50 to 60 nm in diameter, and with surface proteins disposed in an icosahedral-like symmetry (Barrows et al., 2018). Some of these flaviviruses compose a group of arboviruses with global distribution, well known for causing important mosquito and tick-borne diseases in a wide range of vertebrate species, including humans (Gould & Solomon, 2008). As mentioned previously, DENV, WNV, YFV and JEV, as well as multiple other flaviviruses are meaningful human pathogens, generating high morbidity and mortality, and like ZIKV in recent years, others have recently emerged as potential global health threats (Chong et al., 2019).

2.1. Genome organization of flaviviruses

Flaviviruses possess a single-stranded, positive-sense RNA genome of approximately 11 kb (Figure 7) encompassing a single open reading frame (ORF) surrounded by 5' (100 nucleotides) and 3' (400 to 700 nucleotides) untranslated regions (UTRs) (Markoff, 2003). The viral RNA is capped at the 5’-end but is not polyadenylated (Barrows et al., 2018). A single large polyprotein is translated from the viral genome at the host’s endoplasmic reticulum (ER) membrane, and this polyprotein is subsequently cleaved into viral structural and non-structural proteins (Barnard et al., 2021). Three viral structural proteins are generated, including the capsid (C), pre-membrane (prM) and envelope (E) proteins, and they correspond to the main components of the flavivirus virion. Seven viral non-structural proteins (NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5) and one peptide (2k)

are also cleaved from the viral-encoded polyprotein, which are essential for coordinating viral genome replication, transcription, translation and immune evasion (Barnard et al., 2021). All eleven proteins stand in a conserved order among all flavivirus: C–prM–E–NS1–NS2A–NS2B–NS3–NS4A–2k–NS4B–NS5 (Hackett et al., 1985). The structures and functions of all these proteins are well conserved across the *Flaviviridae* family (Pierson & Kielian, 2013). All their prominent roles are summarized in Table 1. However, more recent studies also pointed

towards the possibility that new proteins, encoded by overlapping genes and translated by ribosomal frameshifting, could still play a role in the replication/natural maintenance cycles of flaviviruses. Among them stands the so-called trans-frame fusion protein (designated *fifo*, with around 275 amino acids), and has been associated with the genomes of insect-specific flaviviruses thanks to a frameshifting event in the NS2A coding region (Firth et al., 2010).

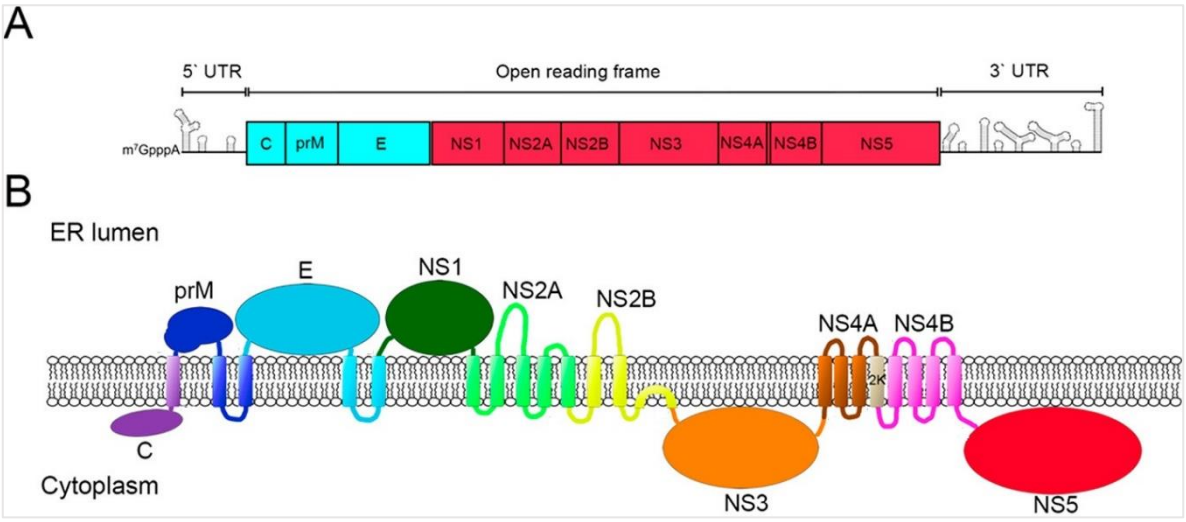


Fig. 7: (A) Representation of the flaviviral genome. (B) Flaviviral polyprotein topology, with predicted transmembrane domains. UTR: untranslated region; ER: endoplasmic reticulum; NS: non-structural. Reprinted with permission from Barrows, N. J., Campos, R. K., Liao, K. C., Prasanth, K. R., Soto-Acosta, R., Yeh, S. C., Garcia-Blanco, M. A. (2018). Biochemistry and Molecular Biology of Flaviviruses. Chemical Reviews, 118(8), 4448–4482. Copyright © American Chemical Society, 2022.

Table 1 – Main known functions of flaviviruses structural and non-structural proteins.

Protein	Function	Reference
C	assembly process	Tan et al., 2020
prM/M	assembly process, secures E protein from pH-induced conformational changes	Roby et al., 2015
E	cell receptor binding and entry	Agrelli et al., 2019
NS1	virus genome replication and immune system evasion	Puerta-Guardo et al., 2019
NS2A	virion assemblage, immune system regulation and evasion	Zhang et al., 2019
NS2B	NS3 cofactor	Luo et al., 2015
NS3	multifunctional enzyme involved in viral genome replication and polyprotein cleavage	Davidson et al., 2020
NS4A	NS3 cofactor, immune system modulation and evasion	Gopala et al., 2018
2k	signal sequence for transfer of NS4B into the endoplasmic reticulum	Roosendaal et al., 2006
NS4B	immune system modulation and evasion	Gopala et al., 2018
NS5	viral genome synthesis, immune system modulation and evasion	Fajardo et al., 2020

2.2. Host range and transmission cycle of flaviviruses

Even though all flaviviruses possess similar genomic organizations, their host range and transmission can be quite different. Most recognized flaviviruses are either considered worldwide health hazards causing millions of infections all over the globe (e.g., DENV and WNV), but for some (e.g., JEV and YFV) their current burden on human health and geographical distribution may prospectively expand in the coming years (Chong et al., 2019). These *bona fide* arboviruses are dual-host flaviviruses that spread horizontally between vertebrates' hosts using hematophagous arthropods (mostly mosquitoes or ticks; Blitvich & Firth, 2015). Dual-host flaviviruses are maintained in an enzootic cycle between a vector and, frequently, either a mammalian or avian amplifying host. Some, such as DENV, have adapted to humans to the point where we have become its mammalian maintenance host in urban settings. Other routes of transmission for these

viruses have been reported, and these include human-to-human transmission of DENV via blood transfusion (Slavov et al., 2019), transmission of WNV via solid organ transplantation (Soto et al., 2022), sexual transmission of ZIKV (Sherley & Ong et al., 2018), transmission of tick-born encephalitis (TBE) by consumption of unpasteurized goat milk and cheese (Brockmann et al., 2018), or transplacental transmission of JEV from an infected mother to her fetus (Chaturvedi et al., 1980). Nevertheless, even in the case of sexual transmission of ZIKV, its contribution to the natural viral maintenance cycles is negligible. However, not all flaviviruses drift between arthropods and vertebrates. As mentioned above, some appear to be restricted to vertebrates (and have apparently lost their ability to replicate in arthropods), while others are insect-specific.

2.3. Insect-specific flaviviruses

The definition of Flavivirus species has been established according to their antigenic properties and vector associations, which include mosquito-borne (MBV), tick-borne (TBV), and no-known vector viruses (NKV, i.e., viruses for which no-invertebrate vector is known) (Kuno et al., 1998). A novel group diverging from other known flaviviruses, named insect-specific flavivirus (ISF), would emerge in 1975 with the detection of CFAV (also the first ISV, as mentioned in chapter 1.2.1.) in *Aedes aegypti* cell cultures (Stollar and Thomas, 1975). The great majority of known ISFs cluster in an monophyletic cluster and would later be described as classical ISF (cISF) or lineage I ISF. However, new studies identified ISFs that did not cluster along with lineage I/cISF in a flavivirus phylogenetic tree. Instead, they appeared to be more closely related to mosquito-borne arboviruses, indicating insect host restriction was not exclusive to cISF (Harrison et al., 2020). These distinct ISF were thought to have eventually lost their ability to replicate in vertebrates (Blitvich and Firth, 2015), and would eventually be renamed dual-host affiliated ISF (dhISF) or lineage II ISF.

cISF can infect and replicate in insect cells but not in vertebrate cells. Experimental studies suggest they seemingly persist in nature primarily by vertical transmission, by which an infected female directly transmits the virus to her progeny (Farfan-Ale et al., 2010; Saiyasombat et al., 2011). Thanks to recent advancements in methods for virus

detection, there has been a sizeable increase in the number of cISFs discovered over the last decade (Blitvich & Firth, 2015).

2.4. History, geographic distribution, and host range of Classical ISF

Classical ISFs have a ubiquitous geographic distribution, with viruses being identified in Europe (Vázquez et al., 2012), Asia (Kyaw Kyaw et al., 2018), America (Gravina et al., 2019), Africa (Villinger et al., 2017) and Australia (McLean et al., 2015). After the initial discovery of CFAV, and as already mentioned above, 20 years passed until the study of ISF was rekindled with the description of the Kamiti river virus in the early 2000s, isolated from *Aedes macintoshi* collected in Kenya in 1999 (Crabtree et al., 2003).

In 2003, the first cISF isolated from *Culex* mosquitoes (from Japan) was discovered (Hoshino et al., 2007), and since then, cISF has been isolated from multiple *Culex* species (Datta et al., 2015; Grisenti et al., 2015; Kyaw Kyaw et al., 2018). Also, several cISF from different mosquito species from multiple genera have been described, including the Nakiwogo virus (in *Mansonia africana* from Uganda; Cook et al., 2009), the Palm Creek virus (from Australian *Coquillettidia xanthogaster*; Hobson-Peters et al., 2013) or the Quang Binh virus (isolated from *Anopheles sinensis* from Vietnam; Crabtree et al., 2009). ISFs have also been detected in the Portuguese continental territory via research work done at IHMT, as previously mentioned in page 9.

While the evolutionary history of cISF is vague and complex, their basal position in phylogenetic trees has led to the suggestion they correspond to an ancestral lineage of flaviviruses (Cook et al., 2012). Their evolutionary history is clouded by the fact that sequences related to extant genomes have been found in the genome of a diverse array of mosquito species as endogenous viral elements (also known as EVEs; Roiz et al., 2009; Crochu et al., 2004; Abílio et al., 2020). Furthermore, cISF research has gained increasing momentum thanks to their potential uses as biological agents, as described in chapter 1.2.3. Multiple research projects, including experiments carried out both *in-vivo* and *in-vitro*, have already demonstrated their capacity to interfere with vector competence via

superinfection exclusion (Kent et al., 2010; Hobson-Peters et al., 2013; Goenaga et al., 2020).

3. The Order *Nidovirales*

The Order *Nidovirales* comprises a genetically distinctive assemblage of enveloped, approximately spherical viruses. They can infect a vast range of hosts, from mammals to insects, and possess the biggest known non-segmented viral RNA genomes, with sizes ranging from 13 to 16 kb for arteriviruses to 26–34 kb for roni- and coronaviruses (Gorbalenya et al., 2006). These RNA viruses are linear single-stranded, positive-sense, capped and polyadenylated (Gorbalenya et al., 2006). According to the ICTV, they are taxonomically (mid-2021) distributed in eight suborders and 14 families (https://talk.ictvonline.org/taxonomy/), including the extensively studied *Arteriviridae* and *Coronaviridae* (Figure 8). The subfamily *Coronavirinae* (in the *Coronaviridae* family) is the one that contains the most significant number of known nidoviruses, including numerous human pathogens, which are then classified into four different genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* (Cong & Ren, 2014). While little is still known about the *Arteriviridae* and *Roniviridae* families, concerns over public health and economic impact of specific pathogenic viruses in both those families, like the yellow head virus (Dong et al., 2017) and the porcine reproductive and respiratory syndrome virus (Ruedas-Torres et al., 2021), have spiked interest in their research. The more recently attested *Mesoniviridae* family is the most under-represented of the families and has only seen their first species described in 2009 (Junglen et al., 2009).

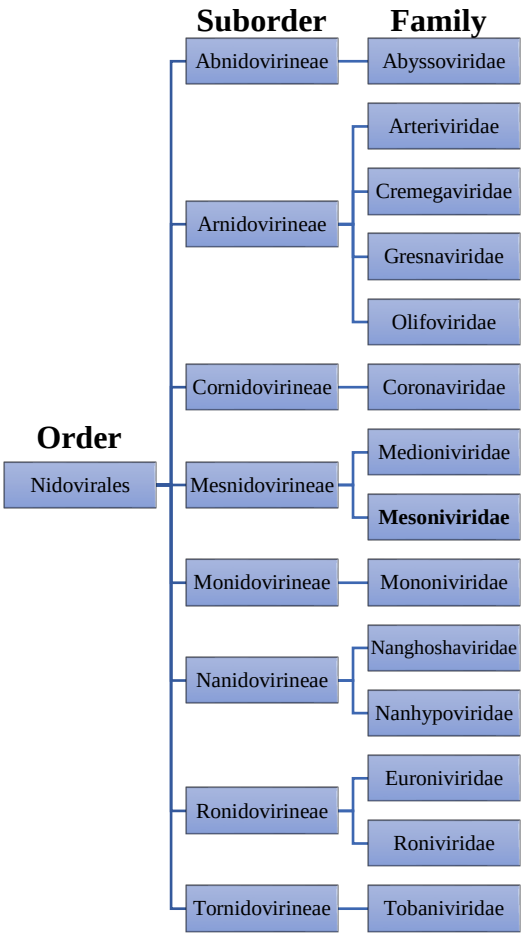


Fig. 8: Taxonomy of the Order *Nidovirales*.

3.1. Mesoniviruses

Mesoniviruses have positive-sense, single-stranded RNA genomes with sizes of ~20 kb and comprise (so far) the only taxon that exclusively consists of insect-specific viruses within the Order *Nidovirales* (Newton et al., 2020). Mesoniviruses are considered insect-specific viruses since no detection or replication has yet been divulged in mammalian hosts (Vasilakis et al., 2014). Typical mesonivirus particles are enveloped, round and their sizes range from 60 to 80 nm in diameter (Zirkel et al., 2013).

3.1.1. History and geographic distribution of mesoniviruses

Mesoniviruses were the first within the Order *Nidovirales* to have been described in 2009 as infectious agents of insects. This description was not, however, detailed, and no

definitive virus-associated formal classification was proposed (Junglen et al., 2009). In fact, their first detailed description was performed in 2011 with the characterization of the Cavally and Nam Dinh viruses, isolated from *Culex* mosquitoes, collected in Cote d'Ivoire and Vietnam, respectively (Nga et al., 2011; Zirkel et al., 2011). Both those studies provided an extensive examination of their genomic sequence and the proteins they encoded, as well as an analysis of their phylogenetic relationship with other nidoviruses. Even though mesoniviruses do not appear to infect vertebrates and, therefore, are not associated with disease in the latter, interest in their research has been steadily rising in the last decade. This has been the result of the curiosity arising from the analysis of their large RNA genome (that encode large numbers of proteins not usually associated with other RNA viruses, including an exoribonuclease (ExoN) involved in the control of replication errors), which may be linked to the evolutionary history of nidoviruses (Lauber et al., 2013), since consecutive increases of ORF1b, ORF1a, and 3′ORFs sizes could be linked to different points in an expansion trajectory of nidovirus genomes. The structural and genetic resemblances of mesoniviruses to other members of the other three predominant families in the Order *Nidovirales* have also contributed to the increase in their research (Vasilakis et al., 2014).

Many different sequences have since been identified and characterized up to the present day, starting in 2012 with the Hana, Meno and Nse viruses, primarily found in *Culex* mosquitoes from Cote d'Ivoire (Zirkel et al., 2013). However, no concrete species demarcation criteria for mesoniviruses existed until 2014, when Vasilakis et al. not only defined 96.8% of amino acid sequence identity as the limit to define new mesonivirus viral species (using RNA-dependent RNA polymerase sequences) but also further characterized two new species, Karang Sari and Kamphaeng Phet. Since then, multiple mesonivirus sequences have been identified from mosquitoes collected in the Americas (Kadiweu and Ofaie virus – Pauvolid-Corrêa et al., 2016; Houston virus – Charles et al., 2018), Africa (Odorna virus – Amoa-Bosompem et al., 2020; Dianke virus – Diagne et al., 2020), Australia (Casuarina – Warrilow et al., 2014; Ngewontan virus – Shi et al., 2017) and Asia (Dak Nong – Kuwata et al., 2013; Bontang Baru virus – Sadeghi et al., 2017). Even though mesoniviruses seem to have a ubiquitous distribution, only recently have they been described in Europe, with scarce reports describing the detection of *Alphamesonivirus* sequences from natural mosquito populations in France in 2017 (Gil et

al., 2017) and in Spain in 2020 (Birnberg et al., 2020). Also, despite the description at IHMT of a mesonivirus sequence obtained from mosquitoes collected in Portugal (R. Parreira, personal communication), their presence in wild caught specimens was not confirmed in a recent viral survey promoted within the scope of this thesis project. Overall, to date, mesoniviruses are classified into one single genus (*Alphamesonivirus*), eight subgenera and ten different species.

Until recently, mesoniviruses had only been detected in mosquito hosts. Nonetheless, sequences with genetic characteristics similar to the *bona fide* mesoniviruses, and designated *meson-like viruses*, have also been described from *Aphis citricidus* aphids collected in 2012 in China (Chang et al., 2020), and from *Thrips tabaci* thrips collected in 2018 in Italy (Chiapello et al., 2021). This suggests that the host range of mesoniviruses (or at least of mesonivirus-related sequences) might be greater than what is currently known. Additionally, a *meson-like virus* was also detected in 2020 in Italy from *Leveillula taurica*, a fungal pathogen (accession number MN609866).

3.1.2. Genome organization of mesoniviruses

The coding content of the genomes of mesoniviruses are arranged into multiple ORFs. Genome organization has been consistently described as ORF1a-ORF1b-ORF2a-ORF2b-ORF3a-ORF3b-ORF4 (Figure 9), but exceptions do exist (e.g., the Meno virus does not encode ORF4; Zirkel et al., 2013). A large section on the 5′ half of the genome encodes two polyproteins (ORF1a and ORF1b), with ORF1b being translated as a fusion polyprotein to ORF1a by ribosomal frameshift, followed by proteolytic processing (Nga et al., 2011). These two ORFs overlap and encode two polyproteins, pp1a, which is characterized by a 3C-like main protease domain, and pp1ab, from which are excised the RNA-dependent RNA polymerase (RdRp) and other conserved replicase-related products, including a superfamily 1 helicase, the ExoN exoribonuclease, a guanine-N7 methyltransferase (NMT), and a ribose-2'-O-methyltransferase (OMT) (Lauber et al., 2012). The 3′ region of the viral genome includes smaller ORFs that encode structural proteins, including putative spike (ORF2a) and nucleocapsid (ORF2b) proteins, as well as ORF4, which encodes a product of unknown function (Nga et al., 2011). The number

of small ORFs differs among distinct viruses in the Order Nidovirales (Gorbalenya et al., 2006).

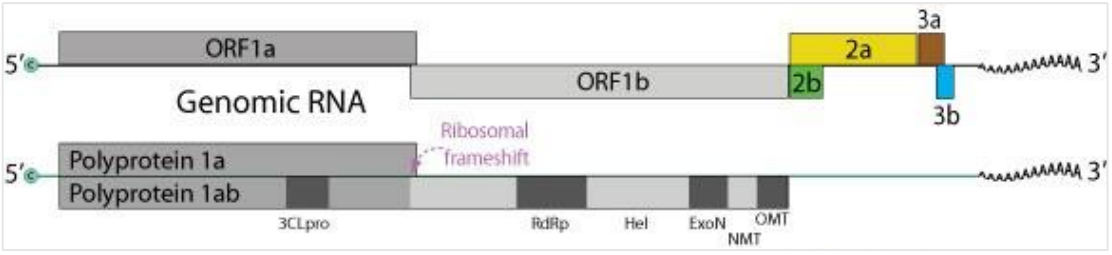


Fig. 9: Representation of most conversed regions of the genome of mesoniviruses, including the ribosomal frameshift responsible for the translation of two polyproteins. 3Clpro: 3C-like protease; ExoN: Exoribonuclease; Hel: Helicase; NMT = N7-methyltransferase; OMT = Nucleoside-2'-O-methyltransferase; ORF: Open reading frame; RdRp: RNA-dependent RNA polymerase. Retrieved from ViralZone, www.expasy.org/viralzone © Swiss Institute of Bioinformatics.

4. The *Parvoviridae* family

The *Parvoviridae* family, one of the viral families many single-stranded DNA (ssDNA) viruses have been assigned to, comprehends small viruses that infect a broad variety of vertebrate and invertebrate species. Evidence of both horizontal transmission (Kelman et al., 2020) and transmission through the germline (Liu et al., 2011) have already been found for parvoviruses. These remarkably diverse viruses are small (23-28 nm), icosahedral-shaped and non-enveloped (Cotmore et al., 2019). A wide range of diseases can be caused by parvoviruses, from acute to chronic, and are usually more severe in animals. Infamous examples include the infections caused by the canine (Mylonakis et al., 2016) and porcine (Mengeling et al., 2000) parvoviruses. On the other hand, human diseases caused by parvoviruses are usually less severe, the two most notable being infections by human parvovirus B19, associated with the "fifth disease" (Weir et al., 2005), and human bocavirus (Guido et al., 2016).

A typical parvovirus genome ranges from 4 to 6 kb, and displays two major expression cassettes (Cotmore et al., 2019). One of these dictates the expression of non-structural (NS) proteins, the largest of which (the so-called non-structural protein 1, or NS1), displays both a highly conserved helicase superfamily domain with helicase and ATPase activity, as well as an endonuclease domain, with site-specific binding activity (Cotmore

et al., 2005). NS1 is also responsible for the induction of cell apoptosis and cell cycle arrest in infected hosts (Lin et al., 2019). Capsid proteins (also simply known as viral proteins or VP) are translated from mRNA transcribed from the second cassette, with the number of structural proteins expressed varying from 1 to 3 between different parvoviruses. These coding regions are flanked by palindromic sequences that form a hairpin-like structure that is essential for replication (Cotmore et al., 2019), and where a host-polymerase recognizes the 5'-end of the viral genome as the primer for replication initiation. There are slight differences in the number and disposition of ORFs in the genomes of different parvovirus, as seen in Figure 10.

ssDNA viruses have been known to integrate into numerous of their hosts' genomes as EVEs, suggesting long-term evolution with them, and have an extensive geographic distribution (Metegnier et al., 2015). Parvovirus genomic sequences have been found throughout the animal kingdom, frequently endogenized into the nuclear genomes of various animals, and with an estimated age of tens of millions of years (Liu et al., 2011).

4.1. History and evolution of parvoviruses taxonomy

Like most ssDNA viruses, the origin of the *Parvoviridae* family could result from ancient recombination events combining non-structural genes from DNA contributors (like bacterial plasmids) and structural genes from RNA viruses (Krupovic, 2013). Curiously, the evolution of many parvoviruses has been associated with frequent recombination events (Leal et al., 2012) and high nucleotide substitution rates (Stamenković et al., 2016).

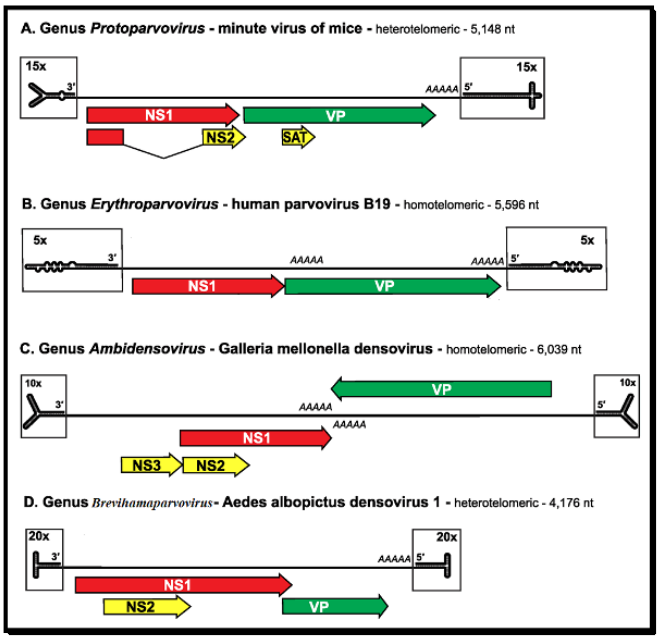


Fig. 10: Representation of genome of viruses from different parvovirus, shown as single lines terminating in boxed hairpin structures (emphasized relative to the rest of the genome). Major open reading frames that encode proteins are displayed as arrowed boxes. NS stands for non-structural, VP stands for viral protein, AAAAA indicates polyadenylation sites, and SAT stands for “small alternatively translated protein”. Retrieved from Cotmore et al. (2019). Reprinted by permission from Microbiology Society under the license number 1202419-1, from 27 Apr 2022.

The family *Parvoviridae* was first established in 1973, but in a taxonomic review dating from 1993, parvoviruses were allocated to either the *Densovirinae* (infecting invertebrates), or the *Parvovirinae* (infecting vertebrates) subfamilies (Cotmore et al., 2019), with subfamily demarcation exclusively supported by the topologies of phylogenetic trees (Muzyczka & Berns, 2001). Initially, subfamily boundaries seemed unlikely to be challenged, as parvoviruses of vertebrates and arthropods had a relatively limited host spectrum. However, as new viruses were discovered and classified, many were assigned to the *Densovirinae* subfamily, where the *Brevidensovirus*, *Penstyldensovirus* and *Hepandensovirus* genus were established.

The growing number of new known viruses successively enlarging the *Densovirinae* subfamily were shown to display higher sequence diversity, departing from the "well-conserved" nature of the pre-established idea of a well conserved *Parvoviridae* family (Cotmore et al., 2014). Therefore, while virus in the *Parvovirinae* subfamily shared high sequence similarity for most of the NS1 protein, that same level of similarity was not

found between viruses from the *Densovirinae* subfamily. In addition, several anomalies impacted the classification of parvoviruses, such as the unexpected isolation of new densoviruses from vertebrates (Yang et al., 2016) and the discovery of ambisense densovirus genomes (genome C in figure 10), adding to the heterogeneous nature of this family. All these events promoted a recent taxonomy revision that took into account both phylogenetic criteria, and amino acid sequence similarity values calculated from comparisons of the sequences either of the whole of the NS1 protein, or strictly considering its helicase domain (Pénzes et al., 2020). This revision led to the split of the *Densovirinae* subfamily into two disparate subfamilies, *Densovirinae* and *Hamaparvovirinae*, with hamaparvoviruses sharing less than 20% amino acid genetic identity of the helicase sequence when compared to other parvoviruses, while sharing between them approximately 30% of NS1 amino acid identity. A group of insect-specific viruses in the *Parvoviridae* family and *Densovirinae* subfamily were previously known as brevidensoviruses. After the latest taxonomy revision, it was discovered they shared about 30% of NS1 protein identity with other hamaparvoviruses. They were, therefore, renamed *Brevihamaparvovirus* and placed into the *Hamaparvovirinae* subfamily (Figure 11; Pénzes et al., 2020).

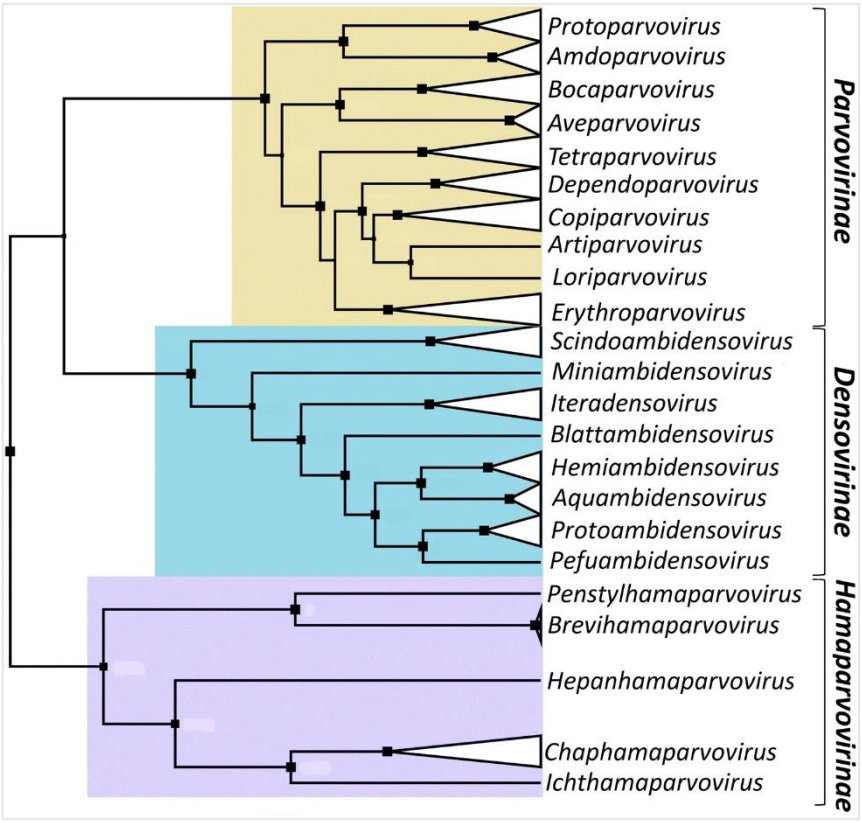


Fig. 11: Phylogenetic representation of relationships at the genera level, based on the Bayesian inference of the helicase domain (167 aa), from which the taxa from the *Parvoviridae* family have recently been established. Retrieved from Pénzes et al. (2020). Reprinted by permission from Springer Nature under the license number 5274360153432, from 22 Mar 2022

4.2. Brevihamaparvoviruses

Unlike other parvoviruses, the brevihamaparvovirus (BHP) host range has been, so far, restricted to a few mosquito species (Pénzes et al., 2020). These viruses are easy to manipulate and have been considered as candidate agents for paratransgenic control of vector-borne diseases (Ren et al., 2014). They have also been reported as able to greatly reduce the susceptibility of *Aedes* mosquitoes to the dengue virus (Wei et al., 2006). However, their research history is relatively recent, with the identification of most BHP from various mosquito cells and wild mosquitoes occurring in the last 30 years.

4.2.1. History and geographic distribution of BHP

The first BHP was isolated from an infected laboratory colony of *Aedes aegypti* in 1972 in Kyiv, in the Ukraine (Lebedeva et al., 1972). Several BHP have been isolated over the years from persistently infected mosquito cell lines, such as the *Haemagogus equinus* cell line GML-HE-12 (O'Neill et al., 1995), the *Aedes albopictus* cell line C6/36 (Chen et al., 2004), the *Anopheles gambiae* cell line Sua5B (Ren et al., 2008) and the *Aedes aegypti* cell line Aag2 (Parry et al., 2019). It is improbable that they evolved from a single contamination event since all these viruses have significant sequence divergence.

The first BHP strain to have been identified in association with wild mosquitoes was described in 1999 from multiple *Aedes aegypti* and *Aedes albopictus* collected in Thailand (Kittayapong et al., 1999). Since then, BHP have been described in different species of *Aedes* and *Culex* from Asia (Zhai et al., 2008) and the Americas (Sadeghi et al., 2017), with singular BHP sequences also being isolated from *Armigeres subalbatus* (Fu et al., 2017) and *Anopheles sinensis* (Zhai et al., 2008). Experimental work in this current thesis project resulted in the first described BHP strains obtained from wild mosquitoes collected in Europe (Portugal, with the first BHP isolation from *Culiseta* mosquitoes) and Africa (Angola, from which they were amplified from mosquitoes from the *Culex pipiens* complex), suggesting a widespread distribution of these parvoviruses.

4.2.2. Genome organization of BHP

The genomes of the members of the *Brevihamaparvovirus* genus possess three ORFs which encode two non-structural proteins (NS1, NS2) and one capsid protein (VP) (Bergoin and Tijssen, 2010). They have some of the smallest ssDNA genomes in the *Parvoviridae* family, with approximately 4 kb. While NS1 has already been described as crucial for initiating viral DNA replication, NS2 participates in suppressing type I interferon responses (Lin et al., 2013) and viral egress from the nucleus, where viral replication occurs. NS1 and NS2 coding sequences overlap, with distinct mRNAs being expressed by alternative splicing, although the mechanisms behind it are still unclear (Chen et al., 2021). The capsid protein is encoded by the VP gene and is vital for viral entry into host cells and the output of infectious viruses (Sánchez-Martínez et al., 2012).

5. Current state of ISV research

As suggested in section 1.2.3, infection of hematophagous arthropods with ISVs could, in a near future, prove as a viable alternative for vector control strategies to reduce mosquito hosts longevity or strategies to reduce replication efficiency of some human pathogenic viruses in their most predominant insect hosts, as well as the permissiveness of infected host cells, by reducing vector competence. However, a positive ecological association has also been reported between ISVs and pathogenic arboviruses, with a report case of multiple co-infections of CxFV and WNV in mosquito pools from Chicago, United States (Newman et al., 2011). As such, the acquisition of genetic and biological information still unknown in multiple ISV groups is essential.

Classical Insect-Specific Flavivirus. Research concerning cISFs is under-represented in the literature as well as in the genetic databases when compared to flaviviruses infecting vertebrates, probably due to their inability to replicate in vertebrate cells, and the fact that they are apathogenic to vertebrates. Therefore, especially when viral surveys involve isolation of viruses using vertebrate cells that are maintained in culture, cISFs could easily have been undetected over the years. Past studies mainly focused on the discovery and genetic characterization of different cISF, as well as their phylogenetic assignment within the genus *Flavivirus*. Additional studies have involved the analysis of their replication in insect cells. However, their origin and spatiotemporal dispersal have rarely been researched, and in the few studies where these have been addressed, no coherent statistical framework was used (Cella et al., 2019). Since little has been done to assess the evolution of cISF over time, we attempted to do so by analyzing either NS5 or complete genome nucleotide sequences of the most representative genetic cISF sublineages. We aimed to genetically characterize the different sublineages of cISF and try to infer their evolutionary history and spatial spread, demonstrating the worth of a Bayesian-based phylodynamic model for the study of ISVs.

Mesonivirus. The only extensive genomic and phylogenetic characterization of mesonivirus, as well as a detailed taxonomy revision of mesoniviruses was performed in 2014 (Vasilakis et al., 2014). At that time, only 13 sequences had been described, and these were assigned to seven species (Alphamesonivirus 1 to Alphamesonivirus 7). However, since then, multiple mesoniviruses have been isolated and characterized, but

no further extensive revisions have been performed. Occasional updates have ensued as new sequences were identified, but none with coherent or detailed presentation and explanation. The latest ICTV update about the taxonomy of the *Mesoniviridae* family was in March 2021 and, unlike in years prior, acknowledges only one single genus (*Alphamesonivirus*), with the addition of eight subgenera, including *Namcalivirus*, represented by the Alphamesonivirus 1 species (consisting of the most substantial number of mesoniviruses isolated to date), and the Alphamesonivirus 10 species (with Dianke virus as the sole representative). Other subgenera, like *Ofalivirus* (Ofaie virus, species Alphamesonivirus 6) or *Casualivirus* (Casuarina virus, species Alphamesonivirus 4), comprehend only one specific viral type. However, recently identified mesoniviruses [e.g., the Odorna virus (OdoV)] remain unclassified.

With scarce detailed reevaluations of mesoniviruses genetic and taxonomy characterization, the recent isolation of multiple mesoniviruses prompted us to reevaluate their position within the family. However, that is not the only factor encouraging a much-needed intensive survey of mesonivirus sequences. Since mesoniviruses are mostly restricted to mosquitoes, and past studies suggested viruses from the Order *Nidovirales* may have evolved in arthropods (Nga et al., 2011), they could hold critical information about the evolution of viruses within the Order *Nidovirales*. Additionally, even if these insect-specific viruses are distantly related to coronaviruses, the current pandemic spread of SARS-CoV-2 coronavirus increased the interest in the study of mesoniviruses as members of a larger group of viruses with overt impact on human health (Lai et al., 2020). Furthermore, the recent discovery of the *meson-like virus* in organisms other than mosquitoes could also hold new information regarding their phylogenetic relationship with other mesoniviruses and viruses from the Order *Nidovirales*.

Brevihamaparvovirus. Even though multiple BHP have been extensively characterized (Ren et al., 2008) and evaluated for their potential for novel genetic strategies to control mosquito vectors (Ren et al., 2014), no consistent description of their most basic genetic traits, such as genetic diversity, recombination, substitution rates, selective pressure acting in the viral genome or phylogenetic and phylodynamic reconstructions, have ever been performed. Also, even though there is already an archaic taxonomic structure for BHP (with two species identified), no evidence or factual data has been presented for classification and demarcation of BHP species, even though demarcation criteria were

already established for parvovirus sequences (Pénzes et al., 2020). Also, as many BHP have been isolated in recent years, multiple BHP sequences remain unclassified. This genera's extensive genetic diversity analysis could help provide new information and complement existing taxonomic classification.

6. Objectives and thesis outline

This thesis entails one main objective - to provide **new information about the biodiversity of ISVs** – as a way to fill the gaps in information that were explored in Chapter 5. This will be done by two specific methods, both independent and complementary, as a way to maximize new information generated in this project:

- **Experimental work:** we sought to obtain new viral sequences for all three ISV groups selected, using as starting material RNA/DNA extracts prepared from acellular macerates prepared from pools of mosquitoes collected during viral screenings carried out in Portugal, Mozambique and Angola. This contribution was carried out by detecting partial genomic sequences by reverse transcription-polymerase chain reaction (RT-PCR) or PCR (to DNA viruses) using both published primers and experimental amplification conditions (*Flavivirus*; Vázquez et al., 2012) as well as those developed in the course of this work (mesonivirus/brevihamaparvovirus).
- **ISV's genetic characterization and analysis of their evolution through both time and space:** this will be accomplished by assembly of multiple datasets for each ISV group, containing both genomic sequences obtained via experimental work (done within the scope of this project) and collection of all publicly available genomic sequences, either partial or complete.

As such, we not only opted to provide new information on these three specific groups of ISVs, but also sought to identify new ISV sequences in mosquito batches collected from different countries. Various bioinformatic tools will be used to execute a genetic characterization of these three groups of viruses and will focus mainly on basic genomic traits: **genetic distance** – quantification of genetic divergence between these groups or between specific populations inside them; **mutation rates** – rates of nucleotide

substitution over time; **selective pressure** – how the genomic composition of a gene will evolve according to the type and number of existing mutation events; **entropy** – measure of the possibility of variation in the information coded by a nucleotide sequence and **recombination events** – exchange of genetic information between two different viruses. We will also measure the quality of information in the produced datasets by analyzing phylogenetic signals via likelihood mapping and temporal signal via root-to-tip linear regression analysis. This will be executed as a way to both analyze and clarify their origin and evolution through both time and space, as well as provide valuable new information regarding genetic diversity, phylogenetic relationships and phylodynamic reconstructions, which may be essential in evaluating their rapport with other arboviruses as well as their potential roles as vector control strategies.

Multiple studies were performed to accomplish these objectives, which are described in the following chapters:

- **Chapters 2 and 3** are devoted to the detection of RNA and DNA viruses, including those from the three target ISVs groups of this thesis, using mosquitoes collected in both Portugal (Chapter 2) and Angola (Chapter 3), in the context of viral surveys conducted between 2015 and 2018.
- **Chapter 4** describes the genetic characterization and spatiotemporal dynamics of cISF. Genomic sequences were investigated with a wide range of molecular tools and we were able to characterize different lineages of cISF. This chapter also provided the reconstruction of the evolutionary history and spatiotemporal dispersal of specific cISF sub-lineages.
- **Chapters 5 and 6** describes similar research conducted in chapter 4 but this time in two other ISV groups, mesoniviruses (Chapter 5) and brevihamaparvoviruses (Chapter 6). An extensive genetic characterization of all available sequences was performed in both cases, accompanied by slight taxonomy revisions. While a robust reconstruction of the evolutionary history of mesoniviruses was not possible, it was performed for brevihamaparvoviruses (albeit with some limitations).
- **Chapter 7** provides additional results not yet explored in former chapters, from a more detailed look into selective pressure and temporal signal analyses to an

attempt to reconstruct the evolutionary history and spatiotemporal dispersal of all cISF.

- **Chapter 8** is devoted to the final remarks, encompassing outlooks on proposed future research studies.

References

Abílio, A. P., Silva, M., Kampango, A., Narciso, I., Gudo, E. S., Das Neves, L. C. B., Parreira, R. (2020). A survey of RNA viruses in mosquitoes from Mozambique reveals novel genetic lineages of flaviviruses and phenuiviruses, as well as frequent flavivirus-like viral DNA forms in Mansonia. BMC Microbiology, 20(1), 225. https://doi.org/10.1186/s12866-020-01905-5

Abílio, A. P., Silva, M., Kampango, A., Narciso, I., Gudo, E. S., Das Neves, L. C. B., Parreira, R. (2020). A survey of RNA viruses in mosquitoes from Mozambique reveals novel genetic lineages of flaviviruses and phenuiviruses, as well as frequent flavivirus-like viral DNA forms in Mansonia. BMC Microbiology, 20(1), 225. https://doi.org/10.1186/s12866-020-01905-5

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S. V., Avšič-Županc, T., Ballinger, M. J., Kuhn, J. H. (2019). Taxonomy of the order Bunyavirales: update 2019. Archives of Virology, 164(7), 1949–1965. https://doi.org/10.1007/s00705-019-04253-6

Achee, N. L., Grieco, J. P., Vatandoost, H., Seixas, G., Pinto, J., Ching-Ng, L., Vontas, J. (2019). Alternative strategies for mosquito-borne arbovirus control. PLoS Neglected Tropical Diseases, 13(1), 1–22. https://doi.org/10.1371/journal.pntd.0006822

Adelman, Z. N., Blair, C. D., Carlson, J. O., Beaty, B. J., & Olson, K. E. (2001). Sindbis virus-induced silencing of dengue viruses in mosquitoes. Insect Molecular Biology, 10(3), 265–273. https://doi.org/10.1046/j.1365-2583.2001.00267.x

Agrelli, A., de Moura, R. R., Crovella, S., & Brandão, L. A. C. (2019). ZIKA virus entry mechanisms in human cells. Infection, Genetics and Evolution, 69(November 2018), 22–29. https://doi.org/10.1016/j.meegid.2019.01.018

Aliota, M. T., Walker, E. C., Uribe Yepes, A., Dario Velez, I., Christensen, B. M., & Osorio, J. E. (2016). The wMel Strain of Wolbachia Reduces Transmission of Chikungunya Virus in Aedes aegypti. PLOS Neglected Tropical Diseases, 10(4), e0004677. https://doi.org/10.1371/journal.pntd.0004677

Amoa-Bosompem, M., Kobayashi, D., Murota, K., Faizah, A. N., Itokawa, K., Fujita, R., Iwanaga, S. (2020). Entomological Assessment of the Status and Risk of Mosquito-borne Arboviral Transmission in Ghana. Viruses, 12(2), 147. https://doi.org/10.3390/v12020147

Bachan, S., & Dinesh-Kumar, S. P. (2012). Tobacco rattle virus (TRV)-based virus-induced gene silencing. Methods in Molecular Biology, 894, 83–92. https://doi.org/10.1007/978-1-61779-882-5_6

Barnard, T. R., Abram, Q. H., Lin, Q. F., Wang, A. B., & Sagan, S. M. (2021). Molecular Determinants of Flavivirus Virion Assembly. Trends in Biochemical Sciences, 46(5), 378–390. https://doi.org/10.1016/j.tibs.2020.12.007

Barrows, N. J., Campos, R. K., Liao, K. C., Prasanth, K. R., Soto-Acosta, R., Yeh, S. C., Garcia-Blanco, M. A. (2018). Biochemistry and Molecular Biology of Flaviviruses. Chemical Reviews, 118(8), 4448–4482. https://doi.org/10.1021/acs.chemrev.7b00719

Barzon, L. (2018). Ongoing and emerging arbovirus threats in Europe. Journal of Clinical Virology, 107, 38–47. https://doi.org/10.1016/j.jcv.2018.08.007

Bergoin, M., & Tijssen, P. (2010). Densoviruses: a highly diverse group of arthropod parvoviruses. Insect Virology, 59–82.

Birnberg, L., Temmam, S., Aranda, C., Correa-Fiz, F., Talavera, S., Bigot, T., Busquets, N. (2020). Viromics on Honey-Baited FTA Cards as a New Tool for the Detection of Circulating Viruses in Mosquitoes. Viruses, 12(3), 274. https://doi.org/10.3390/v12030274

Blitvich, B. J., & Firth, A. E. (2015). Insect-specific flaviviruses: A systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. Viruses, 7(4), 1927–1959. https://doi.org/10.3390/v7041927

Blome, S., Staubach, C., Henke, J., Carlson, J., & Beer, M. (2017). Classical Swine Fever—An Updated Review. Viruses, 9(4), 86. https://doi.org/10.3390/v9040086

Bolling, B. G., Eisen, L., Moore, C. G., & Blair, C. D. (2011). Insect-specific flaviviruses from Culex mosquitoes in Colorado, with evidence of vertical transmission. American Journal of Tropical Medicine and Hygiene, 85(1), 169–177. https://doi.org/10.4269/ajtmh.2011.10-0474

Bolling, B. G., Olea-Popelka, F. J., Eisen, L., Moore, C. G., & 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(2), 90–97. https://doi.org/10.1016/j.virol.2012.02.016

Brockmann, S. O., Oehme, R., Buckenmaier, T., Beer, M., Jeffery-Smith, A., Spannenkrebs, M., Haag-Milz, S., Wagner-Wiening, C., Schlegel, C., Fritz, J., Zange, S., Bestehorn, M., Lindau, A., Hoffmann, D., Tiberi, S., Mackenstedt, U., & Dobler, G. (2018). A cluster of two human cases of tick-borne encephalitis (TBE) transmitted by unpasteurised goat milk and cheese in Germany, May 2016. Eurosurveillance, 23(15). https://doi.org/10.2807/1560-7917.ES.2018.23.15.17-00336

Calisher, C. H., & 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. https://doi.org/10.1146/annurev-ento-020117-043033

Call, L., Nayfach, S., & Kyrpides, N. C. (2021). Illuminating the Virosphere Through Global Metagenomics. Annual Review of Biomedical Data Science, 4(1), 369–391. https://doi.org/10.1146/annurev-biodatasci-012221-095114

Calzolari, M., Zé-Zé, L., Vázquez, A., Sánchez Seco, M. P., Amaro, F., & Dottori, M. (2016). Insect-specific flaviviruses, a worldwide widespread group of viruses only detected in insects. Infection, Genetics and Evolution, 40, 381–388. https://doi.org/10.1016/j.meegid.2015.07.032

Cammisa-Parks, H., Cisar, L. A., Kane, A., & Stollar, V. (1992). The complete nucleotide sequence of cell fusing agent (CFA): Homology between the nonstructural

proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. Virology, 189(2), 511–524. https://doi.org/10.1016/0042-6822(92)90575-A

Cann, A. J. (2015). Particules. In Principles of Molecular Virology, Sixth Edition (pp. 27-57). Academic Press.

Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., Parreira, R. (2015). Negeviruses found in multiple species of mosquitoes from southern Portugal: Isolation, genetic diversity, and replication in insect cell culture. Virology, 483, 318–328. https://doi.org/10.1016/j.virol.2015.04.021

Carvalho, V. L., & Long, M. T. (2021). Insect-Specific Viruses: An overview and their relationship to arboviruses of concern to humans and animals. Virology, 557, 34–43. https://doi.org/10.1016/j.virol.2021.01.007

Casimiro-Soriguer, C. S., Perez-Florido, J., Fernandez-Rueda, J. L., Pedrosa-Corral, I., Guillot-Sulay, V., Lorusso, N., Sanbonmatsu-Gámez, S. (2021). Phylogenetic Analysis of the 2020 West Nile Virus (WNV) Outbreak in Andalusia (Spain). Viruses, 13(5), 836. https://doi.org/10.3390/v13050836

Cella, E., Benvenuto, D., Donati, D., Garilli, F., Angeletti, S., Pascarella, S., & Ciccozzi, M. (2019). Phylogeny of Culex theileri virus flavivirus in Spain, Myanmar, Portugal and Turkey. Asian Pacific Journal of Tropical Medicine, 12(5), 216–223. https://doi.org/10.4103/1995-7645.259242

Chang, T., Guo, M., Zhang, W., Niu, J., & Wang, J. J. (2020). First Report of a Mesonivirus and Its Derived Small RNAs in an Aphid Species Aphis citricidus (Hemiptera: Aphididae), Implying Viral Infection Activity. Journal of Insect Science (Online), 20(2), 1–4. https://doi.org/10.1093/jisesa/ieaa022

Charles, J., Tangudu, C. S., Hurt, S. L., Tumescheit, C., Firth, A. E., Garcia-Rejon, J. E., Blitvich, B. J. (2018). Detection of novel and recognized RNA viruses in mosquitoes from the Yucatan Peninsula of Mexico using metagenomics and characterization of their in vitro host ranges. Journal of General Virology, 99(12), 1729–1738. https://doi.org/10.1099/jgv.0.001165

Chaturvedi, U. C., Mathur, A., Chandra, A., Das, S. K., Tandon, H. O., & Singh, U. K. (1980). Transplacental infection with Japanese encephalitis virus. Journal of Infectious Diseases, 141(6), 712–715. https://doi.org/10.1093/infdis/141.6.712

Chen, S., Cheng, L., Zhang, Q., Lin, W., Lu, X., Brannan, J., Zhang, J. (2004). Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected Aedes albopictus C6/36 cell line. Virology, 318(1), 123–133. https://doi.org/10.1016/j.virol.2003.09.013

Chen, S., Miao, B., Chen, N., Chen, C., Shao, T., Zhang, X., Tong, D. (2021). SYNCRIP facilitates porcine parvovirus viral DNA replication through the alternative splicing of NS1 mRNA to promote NS2 mRNA formation. Veterinary Research, 52(1), 1–15. https://doi.org/10.1186/s13567-021-00938-6

Chiapello, M., Bosco, L., Ciuffo, M., Ottati, S., Salem, N., Rosa, C., Tavella, L., & Turina, M. (2021). Complexity and Local Specificity of the Virome Associated with Tospovirus-Transmitting Thrips Species. Journal of Virology, 95(21). https://doi.org/10.1128/jvi.00597-21

Chong, H. Y., Leow, C. Y., Abdul Majeed, A. B., & Leow, C. H. (2019). Flavivirus infection—A review of immunopathogenesis, immunological response, and immunodiagnosis. Virus Research, 274(October), 197770. https://doi.org/10.1016/j.virusres.2019.197770

Colmant, A. M. G., Hobson-Peters, J., Bielefeldt-Ohmann, H., van den Hurk, A. F., Hall-Mendelin, S., Chow, W. K., Hall, R. A. (2017). A New Clade of Insect-Specific Flaviviruses from Australian Anopheles Mosquitoes Displays Species-Specific Host Restriction. MSphere, 2(4). https://doi.org/10.1128/msphere.00262-17

Cong, Y., & Ren, X. (2014). Coronavirus entry and release in polarized epithelial cells: a review. Reviews in Medical Virology, 24(5), 308–315. https://doi.org/10.1002/rmv.1792

Cook, S., Moureau, G., Harbach, R. E., Mukwaya, L., Goodger, K., Ssenfuka, F., de Lamballerie, X. (2009). Isolation of a novel species of flavivirus and a new strain of Culex flavivirus (Flaviviridae) from a natural mosquito population in Uganda. Journal of General Virology, 90(11), 2669–2678. https://doi.org/10.1099/vir.0.014183-0

Cook, S., Moureau, G., Kitchen, A., Gould, E. A., de Lamballerie, X., Holmes, E. C., & Harbach, R. E. (2012). Molecular evolution of the insect-specific flaviviruses. Journal of General Virology, 93(2), 223–234. https://doi.org/10.1099/vir.0.036525-0

Cook, S., Chung, B. Y. W., Bass, D., Moureau, G., Tang, S., McAlister, E., Firth, A. E. (2013). Novel Virus Discovery and Genome Reconstruction from Field RNA Samples Reveals Highly Divergent Viruses in Dipteran Hosts. PLoS ONE, 8(11), e80720. https://doi.org/10.1371/journal.pone.0080720

Cotmore, S., & Tattersall, P. (2005). A rolling-hairpin strategy. In Parvoviruses (pp. 171–188). https://doi.org/10.1201/b13393-19

Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. A., Mukha, D. V., Pintel, D. J., Qiu, J., Davison, A. J. (2014). The family Parvoviridae. Archives of Virology, 159(5), 1239–1247. https://doi.org/10.1007/s00705-013-1914-1

Cotmore, S. F., Agbandje-McKenna, M., Canuti, M., Chiorini, J. A., Eis-Hubinger, A. M., Hughes, J., Harrach, B. (2019). ICTV virus taxonomy profile: Parvoviridae. Journal of General Virology, 100(3), 367–368. https://doi.org/10.1099/jgv.0.001212

Crabtree, M. B., Sang, R. C., Stollar, V., Dunster, L. M., & Miller, B. R. (2003). Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. Archives of Virology, 148(6), 1095–1118. https://doi.org/10.1007/s00705-003-0019-7

Crabtree, M. B., Nga, P. T., & Miller, B. R. (2009). Isolation and characterization of a new mosquito flavivirus, Quang Binh virus, from Vietnam. Archives of Virology, 154(5), 857–860. https://doi.org/10.1007/s00705-009-0373-1

Crochu, S., Cook, S., Attoui, H., Charrel, R. N., De Chesse, R., Belhouchet, M., de Lamballerie, X. (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. Journal of General Virology, 85(7), 1971–1980. https://doi.org/10.1099/vir.0.79850-0

da Silva Ferreira, R., de Toni Aquino da Cruz, L. C., de Souza, V. J., da Silva Neves, N. A., de Souza, V. C., Filho, L. C. F., Slhessarenko, R. D. (2020). Insect-specific

viruses and arboviruses in adult male culicids from Midwestern Brazil. Infection, Genetics and Evolution, 85(September), 104561. https://doi.org/10.1016/j.meegid.2020.104561

Datta, S., Gopalakrishnan, R., Chatterjee, S., & Veer, V. (2015). Phylogenetic characterization of a novel insect-specific flavivirus detected in a culex pool, collected from Assam, India. Intervirology, 58(3), 149–154. https://doi.org/10.1159/000381901

Davidson, R. B., Hendrix, J., Geiss, B. J., & McCullagh, M. (2020). RNA-Dependent Structures of the RNA-Binding Loop in the Flavivirus NS3 Helicase. The Journal of Physical Chemistry B, 124(12), 2371–2381. https://doi.org/10.1021/acs.jpcb.0c00457

Diagne, M. M., Gaye, A., Ndione, M. H. D., Faye, M., Fall, G., Dieng, I., Sall, A. A. (2020). Dianke virus: A new mesonivirus species isolated from mosquitoes in Eastern Senegal. Virus Research, 275(June 2019), 197802. https://doi.org/10.1016/j.virusres.2019.197802

Diagne, M. M., Ndione, M. H. D., Gaye, A., Barry, M. A., Diallo, D., Diallo, A., Faye, O. (2021). Yellow Fever Outbreak in Eastern Senegal, 2020–2021. Viruses, 13(8), 1475. https://doi.org/10.3390/v13081475

Dong, X., Liu, S., Zhu, L., Wan, X., Liu, Q., Qiu, L., Huang, J. (2017). Complete genome sequence of an isolate of a novel genotype of yellow head virus from Fenneropenaeus chinensis indigenous in China. Archives of Virology, 162(4), 1149–1152. https://doi.org/10.1007/s00705-016-3203-2

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS Biology, 4(5), 699–710. https://doi.org/10.1371/journal.pbio.0040088

Du, J., Zhang, L., Hu, X., Peng, R., Wang, G., Huang, Y., Yin, F. (2021). Phylogenetic Analysis of the Dengue Virus Strains Causing the 2019 Dengue Fever Outbreak in Hainan, China. Virologica Sinica, 36(4), 636–643. https://doi.org/10.1007/s12250-020-00335-x

Egan, A. N. (2006). Phylogenetics of North American Psoraleeae (Leguminosae): Rates and Dates in a Recent, Rapid Radiation. Brigham Young University.

Fajardo, T., Sanford, T. J., Mears, H. V., Jasper, A., Storrie, S., Mansur, D. S., & Sweeney, T. R. (2020). The flavivirus polymerase NS5 regulates translation of viral genomic RNA. Nucleic Acids Research, 48(9), 5081–5093. https://doi.org/10.1093/nar/gkaa242

Farfan-Ale, J. A., Loroño-Pino, M. A., Garcia-Rejon, J. E., Soto, V., Lin, M., Staley, M., Blitvich, B. J. (2010). Detection of flaviviruses and orthobunyaviruses in mosquitoes in the Yucatan Peninsula of Mexico in 2008. Vector Borne and Zoonotic Diseases (Larchmont, N.Y.), 10(8), 777–783. https://doi.org/10.1089/vbz.2009.0196

Ferreira, D. D., Cook, S., Lopes, Â., De Matos, A. P., Esteves, A., Abecasis, A., Parreira, R. (2013). Characterization of an insect-specific flavivirus (OCFVPT) co-isolated from Ochlerotatus caspius collected in southern Portugal along with a putative new Negev-like virus. Virus Genes, 47(3), 532–545. https://doi.org/10.1007/s11262-013-0960-9

Firth, A. E., Blitvich, B. J., Wills, N. M., Miller, C. L., & Atkins, J. F. (2010). Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. Virology, 399(1), 153–166. https://doi.org/10.1016/j.virol.2009.12.033

Frost, S. D. W., & Volz, E. M. (2013). Modelling tree shape and structure in viral phylodynamics. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1614), 20120208. https://doi.org/10.1098/rstb.2012.0208

Frost, S. D. W., Pybus, O. G., Gog, J. R., Viboud, C., Bonhoeffer, S., & Bedford, T. (2015). Eight challenges in phylodynamic inference. Epidemics, 10, 88–92. https://doi.org/10.1016/j.epidem.2014.09.001

Fu, S., Song, S., Liu, H., Li, Y., Li, X., Gao, X., Liang, G. (2017). ZIKA virus isolated from mosquitoes: a field and laboratory investigation in China, 2016. Science China Life Sciences, 60(12), 1364–1371. https://doi.org/10.1007/s11427-017-9196-8

Fu, Y., Cao, M., Wang, H., Du, Z., Liu, Y., & Wang, X. (2020). Discovery and characterization of a novel insect-specific reovirus isolated from Psammotettix alienus. Journal of General Virology, 101(8), 884–892. https://doi.org/10.1099/jgv.0.001442

Fujita, R., Kato, F., Kobayashi, D., Murota, K., Takasaki, T., Tajima, S., Sawabe, K. (2018). Persistent viruses in mosquito cultured cell line suppress multiplication of flaviviruses. Heliyon, 4(8). https://doi.org/10.1016/j.heliyon.2018.e00736

Geoghegan, J. L., & Holmes, E. C. (2017). Predicting virus emergence amid evolutionary noise. Open Biology, 7(10), 170189. https://doi.org/10.1098/rsob.170189

Gil, P., Rakotoarivony, I., Etienne, L., Albane, M., Benoit, F., Grégory, L., Busquets, N., Birnberg, L., Talavera, S., Aranda, C., et al. First Detection of a Mesonivirus in Culex pipiens in Five Countries Around the Mediterranean Basin. Abstract. In Proceedings of the EPIZONE—11th Annual Meeting ANSES, Paris, France, 19–21 September 2017

Goenaga, S., Kenney, J. L., Duggal, N. K., Delorey, M., Ebel, G. D., Zhang, B., 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(11), 5801–5812. https://doi.org/10.3390/v7112911

Goenaga, S., Goenaga, J., Boaglio, E. R., Enria, D. A., & Levis, S. D. C. (2020). Superinfection exclusion studies using West Nile virus and Culex flavivirus strains from Argentina. Memórias Do Instituto Oswaldo Cruz, 115(5), 1–5. https://doi.org/10.1590/0074-02760200012

Gopala, S. B., Chin, W. X., & Shivananju, N. S. (2018). Dengue virus NS2 and NS4: Minor proteins, mammoth roles. Biochemical Pharmacology, 154(March), 54–63. https://doi.org/10.1016/j.bcp.2018.04.008

Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., & Snijder, E. J. (2006). Nidovirales: Evolving the largest RNA virus genome. Virus Research, 117(1), 17–37. https://doi.org/10.1016/j.virusres.2006.01.017

Gorbalenya, A. E., Krupovic, M., Mushegian, A., Kropinski, A. M., Siddell, S. G., Varsani, A., Adams, M. J., Davison, A. J., Dutilh, B. E., Harrach, B., Harrison, R. L.,

Junglen, S., King, A. M. Q., Knowles, N. J., Lefkowitz, E. J., Nibert, M. L., Rubino, L., Sabanadzovic, S., Sanfaçon, H., … Kuhn, J. H. (2020). The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nature Microbiology, 5(5), 668–674. https://doi.org/10.1038/s41564-020-0709-x

Gould, E., & Solomon, T. (2008). Pathogenic flaviviruses. The Lancet, 371(9611), 500–509. https://doi.org/10.1016/S0140-6736(08)60238-X

Gould, E., Pettersson, J., Higgs, S., Charrel, R., & de Lamballerie, X. (2017). Emerging arboviruses: Why today? One Health, 4, 1–13. https://doi.org/10.1016/j.onehlt.2017.06.001

Gravina, H. D., Suzukawa, A. A., Zanluca, C., Cardozo Segovia, F. M., Tschá, M. K., Martins da Silva, A., Duarte dos Santos, C. N. (2019). Identification of insect-specific flaviviruses in areas of Brazil and Paraguay experiencing endemic arbovirus transmission and the description of a novel flavivirus infecting Sabethes belisarioi. Virology, 527(September 2018), 98–106. https://doi.org/10.1016/j.virol.2018.11.008

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. Science, 303(5656), 327–332. https://doi.org/10.1126/science.1090727

Grisenti, M., Vázquez, A., Herrero, L., Cuevas, L., Perez-Pastrana, E., Arnoldi, D., Rizzoli, A. (2015). Wide detection of Aedes flavivirus in north-eastern Italy – a European hotspot of emerging mosquitoborne diseases. Journal of General Virology, 96(2), 420–430. https://doi.org/10.1099/vir.0.069625-0

Guido, M., Tumolo, M. R., Verri, T., Romano, A., Serio, F., De Giorgi, M., Zizza, A. (2016). Human bocavirus: Current knowledge and future challenges. World Journal of Gastroenterology, 22(39), 8684. https://doi.org/10.3748/wjg.v22.i39.8684

Hackett, R. H., Setlow, B., Setlow, P., Rice, C. M., Lenches, E. M., Eddy, S. R., Strauss, J. H. (1985). Nucleotide Sequence of Yellow Fever Virus: Implications for Flavivirus Gene Expression and Evolution. Science, 229(4715), 726–733.

Haddow, A. D., Guzman, H., Popov, V. L., Wood, T. G., Widen, S. G., Haddow, A. D., Weaver, S. C. (2013). First isolation of Aedes flavivirus in the Western Hemisphere

and evidence of vertical transmission in the mosquito Aedes (Stegomyia) albopictus (Diptera: Culicidae). Virology, 440(2), 134–139. https://doi.org/10.1016/j.virol.2012.12.008

Hall-Mendelin, S., McLean, B. J., Bielefeldt-Ohmann, H., Hobson-Peters, J., Hall, R. A., & Van Den Hurk, A. F. (2016). The insect-specific Palm Creek virus modulates West Nile virus infection in and transmission by Australian mosquitoes. Parasites and Vectors, 9(1). https://doi.org/10.1186/s13071-016-1683-2

Hamel, R., Phanitchat, T., Wichit, S., Vargas, R. E. M., Jaroenpool, J., Diagne, C. T., Missé, D. (2021). New insights into the biology of the emerging tembusu virus. Pathogens, 10(8). https://doi.org/10.3390/pathogens10081010

Harapan, H., Michie, A., Sasmono, R. T., & Imrie, A. (2020). Dengue: A minireview. Viruses, 12(8). https://doi.org/10.3390/v12080829

Harrison, J. J., Hobson-Peters, J., Colmant, A. M. G., Koh, J., Newton, N. D., Warrilow, D., Hall, R. A. (2020). Antigenic Characterization of New Lineage II Insect-Specific Flaviviruses in Australian Mosquitoes and Identification of Host Restriction Factors. MSphere, 5(3). https://doi.org/10.1128/mSphere.00095-20

Henderson, A. R. (2005). The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. Clinica Chimica Acta, 359(1–2), 1–26. https://doi.org/10.1016/j.cccn.2005.04.002

Hill, V., & Baele, G. (2019). Bayesian Estimation of Past Population Dynamics in BEAST 1.10 Using the Skygrid Coalescent Model. Molecular Biology and Evolution, 36(11), 2620–2628. https://doi.org/10.1093/molbev/msz172

Ho, S. Y. W., & Duchêne, S. (2014). Molecular-clock methods for estimating evolutionary rates and timescales. Molecular Ecology, 23(24), 5947–5965. https://doi.org/10.1111/mec.12953

Hobson-Peters, J., Yam, A. W. Y., Lu, J. W. F., Setoh, Y. X., May, F. J., Kurucz, N., Hall, R. A. (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(2), e56534. https://doi.org/10.1371/journal.pone.0056534

Hobson-Peters, J., Warrilow, D., Mclean, B. J., Watterson, D., Colmant, A. M. G., Hurk, A. F. Van Den, Hall, R. A. (2016). Discovery and characterisation of a new insect-specific bunyavirus from Culex mosquitoes captured in northern Australia. Virology, 489, 269–281. https://doi.org/10.1016/j.virol.2015.11.003

Holder, M., & Lewis, P. O. (2003). Phylogeny estimation: Traditional and Bayesian approaches. Nature Reviews Genetics, 4(4), 275–284. https://doi.org/10.1038/nrg1044

Hoshino, K., Isawa, H., Tsuda, Y., Yano, K., Sasaki, T., Yuda, M., Sawabe, K. (2007). Genetic characterization of a new insect flavivirus isolated from Culex pipiens mosquito in Japan. Virology, 359(2), 405–414. https://doi.org/10.1016/j.virol.2006.09.039

Huntington, M. K., Allison, J., & Nair, D. (2016). Emerging vector-borne diseases. American Family Physician, 94(7), 551–557. PMID: 27929218.

Ibrahim, B., McMahon, D. P., Hufsky, F., Beer, M., Deng, L., Mercier, P. Le, Marz, M. (2018). A new era of virus bioinformatics. Virus Research, 251(April), 86–90. https://doi.org/10.1016/j.virusres.2018.05.009

Indriani, C., Tantowijoyo, W., Rancès, E., Andari, B., Prabowo, E., Yusdi, D., Utarini, A. (2020). Reduced dengue incidence following deployments of Wolbachia-infected Aedes aegypti in Yogyakarta, Indonesia: a quasi-experimental trial using controlled interrupted time series analysis. Gates Open Research, 4, 50. https://doi.org/10.12688/gatesopenres.13122.1

Jiggins, F. M. (2017). The spread of Wolbachia through mosquito populations. PLoS Biology, 15(6), 1–6. https://doi.org/10.1371/journal.pbio.2002780

Junglen, S., Kurth, A., Kuehl, H., Quan, P.-L., Ellerbrok, H., Pauli, G., Leendertz, F. H. (2009). Examining Landscape Factors Influencing Relative Distribution of Mosquito Genera and Frequency of Virus Infection. EcoHealth, 6(2), 239–249. https://doi.org/10.1007/s10393-009-0260-y

Junglen, S., & Drosten, C. (2013). Virus discovery and recent insights into virus diversity in arthropods. Current Opinion in Microbiology, 16(4), 507–513. https://doi.org/10.1016/j.mib.2013.06.005

Junglen, S., Korries, M., Grasse, W., Wieseler, J., Kopp, A., Hermanns, K., Kümmerer, B. M. (2017). Host Range Restriction of Insect-Specific Flaviviruses Occurs at Several Levels of the Viral Life Cycle. MSphere, 2(1), 1–15. https://doi.org/10.1128/mSphere.00375-16

Kauffman, K. M., Hussain, F. A., Yang, J., Arevalo, P., Brown, J. M., Chang, W. K., Polz, M. F. (2018). A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. Nature, 554(7690), 118–122. https://doi.org/10.1038/nature25474

Kelman, M., Harriott, L., Carrai, M., Kwan, E., Ward, M. P., & Barrs, V. R. (2020). Phylogenetic and Geospatial Evidence of Canine Parvovirus Transmission between Wild Dogs and Domestic Dogs at the Urban Fringe in Australia. Viruses, 12(6), 663. https://doi.org/10.3390/v12060663

Kent, R. J., Crabtree, M. B., & Miller, B. R. (2010). Transmission of West Nile Virus by Culex quinquefasciatus Say Infected with Culex Flavivirus Izabal. PLoS Neglected Tropical Diseases, 4(5), e671. https://doi.org/10.1371/journal.pntd.0000671

Kittayapong, P., Baisley, K. J., & O'Neill, S. L. (1999). A mosquito densovirus infecting Aedes aegypti and Aedes albopictus from Thailand. American Journal of Tropical Medicine and Hygiene, 61(4), 612–617. https://doi.org/10.4269/ajtmh.1999.61.612

Koh, C., Henrion-Lacritick, A., Frangeul, L., & Saleh, M. C. (2021). Interactions of the insect-specific palm creek virus with zika and chikungunya viruses in aedes mosquitoes. Microorganisms, 9(8). https://doi.org/10.3390/microorganisms9081652

Krupovic, M. (2013). Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. Current Opinion in Virology, 3(5), 578–586. https://doi.org/10.1016/j.coviro.2013.06.010

Kuno, G., Chang, G.-J. J., Tsuchiya, K. R., Karabatsos, N., & Cropp, C. B. (1998). Phylogeny of the Genus Flavivirus. Journal of Virology, 72(1), 73–83. https://doi.org/10.1128/JVI.72.1.73-83.1998

Kuwata, R., Satho, T., Isawa, H., Yen, N. T., Phong, T. V., Nga, P. T., Sawabe, K. (2013). Characterization of Dak Nong virus, an insect nidovirus isolated from Culex mosquitoes in Vietnam. Archives of Virology, 158(11), 2273–2284. https://doi.org/10.1007/s00705-013-1741-4

Kyaw Kyaw, A., Tun, M. M. N., Buerano, C. C., Nabeshima, T., Sakaguchi, M., Ando, T., Morita, K. (2018). Isolation and genomic characterization of Culex flaviviruses from mosquitoes in Myanmar. Virus Research, 247(January), 120–124. https://doi.org/10.1016/j.virusres.2018.01.007

Lanciotti, R. S., Roehrig, J. T., Deubel, V., Smith, J., Parker, M., Steele, K., Gubler, D. J. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the Northeastern United States. Science, 286(5448), 2333–2337. https://doi.org/10.1126/science.286.5448.2333

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. International Journal of Antimicrobial Agents, 55(3), 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924

Lauber, C., Goeman, J. J., Parquet, M. del C., Thi Nga, P., Snijder, E. J., Morita, K., & Gorbalenya, A. E. (2013). The Footprint of Genome Architecture in the Largest Genome Expansion in RNA Viruses. PLoS Pathogens, 9(7), e1003500. https://doi.org/10.1371/journal.ppat.1003500

Laureti, M., Paradkar, P. N., Fazakerley, J. K., & Rodriguez-Andres, J. (2020). Superinfection exclusion in mosquitoes and its potential as an arbovirus control strategy. Viruses, 12(11). https://doi.org/10.3390/v12111259

Leal, É., Villanova, F. E., Lin, W., Hu, F., Liu, Q., Liu, Y., & Cui, S. (2012). Interclade recombination in porcine parvovirus strains. Journal of General Virology, 93(PART 12), 2692–2704. https://doi.org/10.1099/vir.0.045765-0

Lebedeva, P. O., Zelenko, A. P., Kuznetsova, M. A., & Gudzgorban, A. P. (1972). Studies on the demonstration of a viral infection in larvae of Aedes aegypti mosquitoes. Microbiology Jacksonville State University, 34, 70–73.

Lemey, P., Rambaut, A., Drummond, A. J., & Suchard, M. A. (2009). Bayesian phylogeography finds its roots. PLoS Computational Biology, 5(9). https://doi.org/10.1371/journal.pcbi.1000520

Lemey, P., Rambaut, A., Welch, J. J., & Suchard, M. A. (2010). Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. Molecular Biology and Evolution, 27(8), 1877–1885. https://doi.org/10.1093/molbev/msq067

Lequime, S., & Lambrechts, L. (2017). Discovery of flavivirus-derived endogenous viral elements in Anopheles mosquito genomes supports the existence of Anopheles -associated insect-specific flaviviruses. Virus Evolution, 3(1), vew035. https://doi.org/10.1093/ve/vew035

Li, C. X., Shi, M., Tian, J. H., Lin, X. D., Kang, Y. J., Chen, L. J., Zhang, Y. Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. ELife, 2015(4). https://doi.org/10.7554/eLife.05378

Lin, W., Qiu, Z., Liu, Q., & Cui, S. (2013). Interferon induction and suppression in swine testicle cells by porcine parvovirus and its proteins. Veterinary Microbiology, 163(1–2), 157–161. https://doi.org/10.1016/j.vetmic.2012.11.032

Lin, P., Cheng, Y., Song, S., Qiu, J., Yi, L., Cao, Z., Wang, J. (2019). Viral Nonstructural Protein 1 Induces Mitochondrion-Mediated Apoptosis in Mink Enteritis Virus Infection. Journal of Virology, 93(22). https://doi.org/10.1128/JVI.01249-19

Liu, L., Yu, L., Kubatko, L., Pearl, D. K., & Edwards, S. V. (2009). Coalescent methods for estimating phylogenetic trees. Molecular Phylogenetics and Evolution, 53(1), 320–328. https://doi.org/10.1016/j.ympev.2009.05.033

Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Jiang, D. (2011). Widespread Endogenization of Densoviruses and Parvoviruses in Animal and Human Genomes. Journal of Virology, 85(19), 9863–9876. https://doi.org/10.1128/jvi.00828-11

Luo, D., Vasudevan, S. G., & Lescar, J. (2015). The flavivirus NS2B–NS3 protease–helicase as a target for antiviral drug development. Antiviral Research, 118(September 2005), 148–158. https://doi.org/10.1016/j.antiviral.2015.03.014

Ma, H., Galvin, T. A., Glasner, D. R., Shaheduzzaman, S., & Khan, A. S. (2014). Identification of a Novel Rhabdovirus in Spodoptera frugiperda Cell Lines. Journal of Virology, 88(12), 6576–6585. https://doi.org/10.1128/jvi.00780-14

Madani, T. A., Azhar, E. I., Abuelzein, E. T. M. E., Kao, M., Al-Bar, H. M. S., Abu-Araki, H., Ksiazek, T. G. (2011). Alkhumra (Alkhurma) virus outbreak in Najran, Saudi Arabia: Epidemiological, clinical, and Laboratory characteristics. Journal of Infection, 62(1), 67–76. https://doi.org/10.1016/j.jinf.2010.09.032

Marklewitz, M., Zirkel, F., Rwego, I. B., Heidemann, H., Trippner, P., Kurth, A., Junglen, S. (2013). Discovery of a Unique Novel Clade of Mosquito-Associated Bunyaviruses. Journal of Virology, 87(23), 12850–12865. https://doi.org/10.1128/jvi.01862-13

Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C., & Junglen, S. (2015). Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. Proceedings of the National Academy of Sciences, 112(24), 7536–7541. https://doi.org/10.1073/pnas.1502036112

Markoff, L. (2003). 5′- and 3′-noncoding regions in flavivirus RNA. In Advances in Virus Research (Vol. 59, pp. 177–228). https://doi.org/10.1016/S0065-3527(03)59006-6

Martina, B. E., Barzon, L., Pijlman, G. P., de la Fuente, J., Rizzoli, A., Wammes, L. J., Papa, A. (2017). Human to human transmission of arthropod-borne pathogens. Current Opinion in Virology, 22, 13–21. https://doi.org/10.1016/j.coviro.2016.11.005

Martinez, J., Longdon, B., Bauer, S., Chan, Y. S., Miller, W. J., Bourtzis, K., … Jiggins, F. M. (2014). Symbionts Commonly Provide Broad Spectrum Resistance to Viruses in Insects: A Comparative Analysis of Wolbachia Strains. PLoS Pathogens, 10(9). https://doi.org/10.1371/journal.ppat.1004369

Mengeling, W. L., Lager, K. M., & Vorwald, A. C. (2000). The effect of porcine parvovirus and porcine reproductive and respiratory syndrome virus on porcine reproductive performance. Animal Reproduction Science, 60–61, 199–210. https://doi.org/10.1016/S0378-4320(00)00135-4

Metegnier, G., Becking, T., Chebbi, M. A., Giraud, I., Moumen, B., Schaack, S., Gilbert, C. (2015). Comparative paleovirological analysis of crustaceans identifies multiple widespread viral groups. Mobile DNA, 6(1). https://doi.org/10.1186/s13100-015-0047-3

McLean, B. J., Hobson-Peters, J., Webb, C. E., Watterson, D., Prow, N. A., Nguyen, H. D., Hall, R. A. (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. https://doi.org/10.1016/j.virol.2015.07.021

Mothes, W., Sherer, N. M., Jin, J., & Zhong, P. (2010). Virus Cell-to-Cell Transmission. Journal of Virology, 84(17), 8360–8368. https://doi.org/10.1128/jvi.00443-10

Mylonakis, M., Kalli, I., & Rallis, T. (2016). Canine parvoviral enteritis: an update on the clinical diagnosis, treatment, and prevention. Veterinary Medicine: Research and Reports, Volume 7, 91–100. https://doi.org/10.2147/VMRR.S80971

Nasar, F., Palacios, G., Gorchakov, R. V., Guzman, H., Travassos Da Rosa, A. P., Savji, N., Weaver, S. C. (2012). Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. Proceedings of the National Academy of Sciences of the United States of America, 109(36), 14622–14627. https://doi.org/10.1073/pnas.1204787109

Nascimento, F. F., Reis, M. Dos, & Yang, Z. (2017). A biologist's guide to Bayesian phylogenetic analysis. Nature Ecology and Evolution, 1(10), 1446–1454. https://doi.org/10.1038/s41559-017-0280-x

Newman, C. M., Cerutti, F., Anderson, T. K., Hamer, G. L., Walker, E. D., Kitron, U. D., Goldberg, T. L. (2011). Culex flavivirus and West Nile virus mosquito coinfection and positive ecological association in Chicago, United States. Vector-Borne and Zoonotic Diseases, 11(8), 1099–1105. https://doi.org/10.1089/vbz.2010.0144

Newton, N. D., Colmant, A. M. G., O'Brien, C. A., Ledger, E., Paramitha, D., Bielefeldt-Ohmann, H., Hobson-Peters, J. (2020). Genetic, morphological and antigenic relationships between mesonivirus isolates from australian mosquitoes and evidence for their horizontal transmission. Viruses, 12(10). https://doi.org/10.3390/v12101159

Nga, P. T., Parquet, M. del C., Lauber, C., Parida, M., Nabeshima, T., Yu, F., Gorbalenya, A. E. (2011). Discovery of the First Insect Nidovirus, a Missing Evolutionary Link in the Emergence of the Largest RNA Virus Genomes. PLoS Pathogens, 7(9), e1002215. https://doi.org/10.1371/journal.ppat.1002215

Nouri, S., Matsumura, E. E., Kuo, Y. W., & Falk, B. W. (2018). Insect-specific viruses: from discovery to potential translational applications. Current Opinion in Virology, 33, 33–41. https://doi.org/10.1016/j.coviro.2018.07.006

Oaks R., J., Cobb, A. K., Minin, N., V., & Leaché, D., A. (2019). Marginal Likelihoods in Phylogenetics: A Review of Methods and Applications. Systematic Biology, 68(5), 681–697. https://doi.org/10.1093/sysbio/syz003

Öhlund, P., Lundén, H., & Blomström, A. L. (2019). Insect-specific virus evolution and potential effects on vector competence. Virus Genes, 55(2), 127–137. https://doi.org/10.1007/s11262-018-01629-9

Oldstone, M. B. A. (2019). History of virology. In Encyclopedia of Microbiology, Fourth Edition (pp. 608-612). Elsevier Inc.

O'Neill, S. L., Kittayapong, P., Braig, H. R., Andreadis, T. G., Gonzalez, J. P., & Tesh, R. B. (1995). Insect densoviruses may be widespread in mosquito cell lines. Journal of General Virology, 76(8), 2067–2074. https://doi.org/10.1099/0022-1317-76-8-2067

Parreira, R., Cook, S., Lopes, Â., de Matos, A. P., de Almeida, A. P. G., Piedade, J., & Esteves, A. (2012). Genetic characterization of an insect-specific flavivirus isolated from Culex theileri mosquitoes collected in southern Portugal. Virus Research, 167(2), 152–161. https://doi.org/10.1016/j.virusres.2012.04.010

Parry, R., Bishop, C., Hayr, L. De, & Asgari, S. (2019). Density-dependent enhanced replication of a densovirus in Wolbachia - infected Aedes cells is associated

with production of piRNAs and higher virus-derived siRNAs. Virology, 528(December 2018), 89–100. https://doi.org/10.1016/j.virol.2018.12.006

Pauvolid-Corrêa, A., Solberg, O., Couto-Lima, D., Nogueira, R. M., Langevin, S., & Komar, N. (2016). Novel viruses isolated from mosquitoes in Pantanal, Brazil. Genome Announcements, 4(6), 5–6. https://doi.org/10.1128/genomeA.01195-16

Pénzes, J. J., Söderlund-Venermo, M., Canuti, M., Eis-Hübinger, A. M., Hughes, J., Cotmore, S. F., & Harrach, B. (2020). Reorganizing the family Parvoviridae: a revised taxonomy independent of the canonical approach based on host association. Archives of Virology, 165(9), 2133–2146. https://doi.org/10.1007/s00705-020-04632-4

Pielnaa, P., Al-Saadawe, M., Saro, A., Dama, M. F., Zhou, M., Huang, Y., Xia, Z. (2020). Zika virus-spread, epidemiology, genome, transmission cycle, clinical manifestation, associated challenges, vaccine and antiviral drug development. Virology, 543, 34–42. https://doi.org/10.1016/j.virol.2020.01.015

Pierson, T. C., & Kielian, M. (2013). Flaviviruses: braking the entering. Current Opinion in Virology, 3(1), 3–12. https://doi.org/10.1016/j.coviro.2012.12.001

Puerta-Guardo, H., Glasner, D. R., Espinosa, D. A., Biering, S. B., Patana, M., Ratnasiri, K., Harris, E. (2019). Flavivirus NS1 Triggers Tissue-Specific Vascular Endothelial Dysfunction Reflecting Disease Tropism. Cell Reports, 26(6), 1598-1613.e8. https://doi.org/10.1016/j.celrep.2019.01.036

Pybus, O. G., & Rambaut, A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. Nature Reviews Genetics, 10(8), 540–550. https://doi.org/10.1038/nrg2583

Qin, X. C., Shi, M., Tian, J. H., Lin, X. D., Gao, D. Y., He, J. R., Zhang, Y. Z. (2014). A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. Proceedings of the National Academy of Sciences of the United States of America, 111(18), 6744–6749.

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution, 2(1), vew007. https://doi.org/10.1093/ve/vew007

Ren, X., Hoiczyk, E., & Rasgon, J. L. (2008). Viral paratransgenesis in the malaria vector Anopheles gambiae. PLoS Pathogens, 4(8), 4–11. https://doi.org/10.1371/journal.ppat.1000135

Ren, X., Hughes, G. L., Niu, G., Suzuki, Y., & Rasgon, J. L. (2014). Anopheles gambiae densovirus (AgDNV) has negligible effects on adult survival and transcriptome of its mosquito host. PeerJ, 2014(1), 1–11. https://doi.org/10.7717/peerj.584

Rife, B. D., Mavian, C., Chen, X., Ciccozzi, M., Salemi, M., Min, J., & Prosperi, M. C. (2017). Phylodynamic applications in 21st century global infectious disease research. Global Health Research and Policy, 2(1), 13. https://doi.org/10.1186/s41256-017-0034-y

Roby, J. A., Setoh, Y. X., Hall, R. A., & Khromykh, A. A. (2015). Post-translational regulation and modifications of flavivirus structural proteins. Journal of General Virology, 96(7), 1551–1569. https://doi.org/10.1099/vir.0.000097

Roiz, D., Vázquez, A., Seco, M. P. S., Tenorio, A., & Rizzoli, A. (2009). Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. Virology Journal, 6(1), 93. https://doi.org/10.1186/1743-422X-6-93

Romo, H., Kenney, J. L., Blitvich, B. J., & 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), 1–13. https://doi.org/10.1038/s41426-018-0180-4

Roosendaal, J., Westaway, E. G., Khromykh, A., & Mackenzie, J. M. (2006). Regulated Cleavages at the West Nile Virus NS4A-2K-NS4B Junctions Play a Major Role in Rearranging Cytoplasmic Membranes and Golgi Trafficking of the NS4A Protein. Journal of Virology, 80(9), 4623–4632. https://doi.org/10.1128/jvi.80.9.4623-4632.2006

Ross, P. A., Turelli, M., & Hoffmann, A. A. (2019). Evolutionary Ecology of Wolbachia Releases for Disease Control. Annual Review of Genetics, 53(1), 93–116. https://doi.org/10.1146/annurev-genet-112618-043609

Ruedas-Torres, I., Rodríguez-Gómez, I. M., Sánchez-Carvajal, J. M., Larenas-Muñoz, F., Pallarés, F. J., Carrasco, L., & Gómez-Laguna, J. (2021). The jigsaw of PRRSV virulence. Veterinary Microbiology, 260, 109168. https://doi.org/10.1016/j.vetmic.2021.109168

Sadeghi, M., Popov, V., Guzman, H., Phan, T. G., Vasilakis, N., Tesh, R., & Delwart, E. (2017). Genomes of viral isolates derived from different mosquitos species. Virus Research, 242, 49–57. https://doi.org/10.1016/j.virusres.2017.08.012

Saiyasombat, R., Bolling, B. G., Brault, A. C., Bartholomay, L. C., & Blitvich, B. J. (2011). Evidence of efficient transovarial transmission of culex flavivirus by Culex pipiens (Diptera: Culicidae). Journal of Medical Entomology, 48(5), 1031–1038. https://doi.org/10.1603/ME11043

Sanjuán, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. Cellular and Molecular Life Sciences, 73(23), 4433–4448. https://doi.org/10.1007/s00018-016-2299-6

Sánchez-Martínez, C., Grueso, E., Carroll, M., Rommelaere, J., & Almendral, J. M. (2012). Essential role of the unordered VP2 n-terminal domain of the parvovirus MVM capsid in nuclear assembly and endosomal enlargement of the virion fivefold channel for cell entry. Virology, 432(1), 45–56. https://doi.org/10.1016/j.virol.2012.05.025

Sherley, M., & Ong, C.-W. (2018). Sexual transmission of Zika virus: a literature review. Sexual Health, 15(3), 183. https://doi.org/10.1071/SH17046

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. Nature, 540(7634), 539–543. https://doi.org/10.1038/nature20167

Shi, M., Neville, P., Nicholson, J., Eden, J.-S., Imrie, A., & Holmes, E. C. (2017). High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-

Associated RNA Viruses in Western Australia. Journal of Virology, 91(17), 1–17. https://doi.org/10.1128/jvi.00680-17

Siddell, S. G., Walker, P. J., Lefkowitz, E. J., Mushegian, A. R., Adams, M. J., Dutilh, B. E., Gorbalenya, A. E., Harrach, B., Harrison, R. L., Junglen, S., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Nibert, M., Rubino, L., Sabanadzovic, S., Sanfaçon, H., Simmonds, P., Davison, A. J. (2019). Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). Archives of Virology, 164(3), 943–946. https://doi.org/10.1007/s00705-018-04136-2

Silva, M., Morais, P., Maia, C., de Sousa, C. B., de Almeida, A. P. G., & Parreira, R. (2019). A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal. Virus Research, 274(July), 197769. https://doi.org/10.1016/j.virusres.2019.197769

Simmonds, P., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Stapleton, J. T. (2017). ICTV Virus Taxonomy Profile: Flaviviridae. Journal of General Virology, 98(1), 2–3. https://doi.org/10.1099/jgv.0.000672

Slavov, S. N., Cilião-Alves, D. C., Gonzaga, F. A. C., Moura, D. R., de Moura, A. C. A. M., de Noronha, L. A. G., Haddad, R. (2019). Dengue seroprevalence among asymptomatic blood donors during an epidemic outbreak in Central-West Brazil. PLoS ONE, 14(3), e0213793. https://doi.org/10.1371/journal.pone.0213793

Smith, D. B., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Simmonds, P. (2016). Proposed update to the taxonomy of the genera Hepacivirus and Pegivirus within the Flaviviridae family. Journal of General Virology, 97(11), 2894–2907. https://doi.org/10.1099/jgv.0.000612

Smith, D. B., Meyers, G., Bukh, J., Gould, E. A., Monath, T., Muerhoff, A. S., Becher, P. (2017). Proposed revision to the taxonomy of the genus Pestivirus, family Flaviviridae. Journal of General Virology, 98(8), 2106–2112. https://doi.org/10.1099/jgv.0.000873

Soto, R. A., McDonald, E., Annambhotla, P., Velez, J. O., Laven, J., Panella, A. J., Gould, C. V. (2022). West Nile Virus Transmission by Solid Organ Transplantation

and Considerations for Organ Donor Screening Practices, United States. Emerging Infectious Diseases, 28(2), 403–406. https://doi.org/10.3201/eid2802.211697

Stamenković, G. G., Ćirković, V. S., Šiljić, M. M., Blagojević, J. V., Knežević, A. M., Joksić, I. D., & Stanojević, M. P. (2016). Substitution rate and natural selection in parvovirus B19. Scientific Reports, 6(October), 1–9. https://doi.org/10.1038/srep35759

Stollar, V., & Thomas, V. L. (1975). An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. Virology, 64(2), 367–377. https://doi.org/10.1016/0042-6822(75)90113-0

Tan, T. Y., Fibriansah, G., Kostyuchenko, V. A., Ng, T. S., Lim, X. X., Zhang, S., Lok, S. M. (2020). Capsid protein structure in Zika virus reveals the flavivirus assembly process. Nature Communications, 11(1), 895. https://doi.org/10.1038/s41467-020-14647-9

Talavera, S., Birnberg, L., Nuñez, A. I., Muñoz-Muñoz, F., Vázquez, A., & Busquets, N. (2018). Culex flavivirus infection in a Culex pipiens mosquito colony and its effects on vector competence for Rift Valley fever phlebovirus. Parasites and Vectors, 11(1). https://doi.org/10.1186/s13071-018-2887-4

Tree, M. O., McKellar, D. R., Kieft, K. J., Watson, A. M., Ryman, K. D., & Conway, M. J. (2016). Insect-specific flavivirus infection is restricted by innate immunity in the vertebrate host. Virology, 497, 81–91. https://doi.org/10.1016/j.virol.2016.07.005

Vasilakis, N., Guzman, H., Firth, C., Forrester, N. L., Widen, S. G., Wood, T. G., Tesh, R. B. (2014). Mesoniviruses are mosquito-specific viruses with extensive geographic distribution and host range. Virology Journal, 11(1), 97. https://doi.org/10.1186/1743-422X-11-97

Vázquez, A., Sánchez-Seco, M. P., Palacios, G., Molero, F., Reyes, N., Ruiz, S., Tenorio, A. (2012). Novel flaviviruses detected in different species of mosquitoes in Spain. Vector-Borne and Zoonotic Diseases, 12(3), 223–229. https://doi.org/10.1089/vbz.2011.0687

Villinger, J., Mbaya, M. K., Ouso, D., Kipanga, P. N., Lutomiah, J., & Masiga, D. K. (2017). Arbovirus and insect-specific virus discovery in Kenya by novel six genera

multiplex high-resolution melting analysis. Molecular Ecology Resources, 17(3), 466–480. https://doi.org/10.1111/1755-0998.12584

ViralZone. (2011). Mesoniviridae. Retrieved November 23, 2021, from: https://viralzone.expasy.org/4776

Wagner, E. K., Hewlett, M. J., Bloom, D. C., & Camerini, D. (2007a). Introduction- The Impact of Viruses on Our View of Life. In Basic Virology, 3rd Edition (pp. 3–14). Wiley-Blackwell.

Wagner, E. K., Hewlett, M. J., Bloom, D. C., & Camerini, D. (2007b). Viruses: New Approaches and New Problems. In Basic Virology, 3rd Edition (pp. 433–499). Wiley-Blackwell.

Wang, Y., Xia, H., Zhang, B., Liu, X., & Yuan, Z. (2017). Isolation and characterization of a novel mesonivirus from Culex mosquitoes in China. Virus Research, 240(June), 130–139. https://doi.org/10.1016/j.virusres.2017.08.001

Warrilow, D., Watterson, D., Hall, R. A., Davis, S. S., Weir, R., Kurucz, N., Hobson-Peters, J. (2014). A new species of Mesonivirus from the Northern Territory, Australia. PLoS ONE, 9(3). https://doi.org/10.1371/journal.pone.0091103

Weaver, S. C., & Barrett, A. D. T. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. Nature Reviews Microbiology, 2(10), 789–801. https://doi.org/10.1038/nrmicro1006

Weaver, S. C., Charlier, C., Vasilakis, N., & Lecuit, M. (2018). Zika, Chikungunya, and Other Emerging Vector-Borne Viral Diseases. Annual Review of Medicine, 69(1), 395–408. https://doi.org/10.1146/annurev-med-050715-105122

Wei, W., Shao, D., Huang, X., Li, J., Chen, H., Zhang, Q., & Zhang, J. (2006). The pathogenicity of mosquito densovirus (C6/36DNV) and its interaction with dengue virus type II in Aedes albopictus. American Journal of Tropical Medicine and Hygiene, 75(6), 1118–1126. https://doi.org/10.4269/ajtmh.2006.75.1118

Weir, E. (2005). Parvovirus B19 infection: fifth disease and more. Canadian Medical Association Journal, 172(6), 743–743. https://doi.org/10.1503/cmaj.045293

World Health Organization. Vector-borne diseases. March 2020. http://www.who.int/mediacentre/factsheets/fs387/en/. Accessed November 09, 2021.

Yang, W. T., Shi, S. H., Jiang, Y. L., Zhao, L., Chen, H. L., Huang, K. Y., Wang, C. F. (2016). Genetic characterization of a densovirus isolated from great tit (Parus major) in China. Infection, Genetics and Evolution, 41, 107–112. https://doi.org/10.1016/j.meegid.2016.03.035

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: Principles and practice. Nature Reviews Genetics, 13(5), 303–314. https://doi.org/10.1038/nrg3186

Young, P. R. (2018). Arboviruses: A family on the move. In Advances in Experimental Medicine and Biology (Vol. 1062, pp. 1–10). https://doi.org/10.1007/978-981-10-8727-1_1

Zhai, Y., Lv, X., Sun, X., Fu, S., Gong, Z., Fen, Y., Liang, G. (2008). Isolation and characterization of the full coding sequence of a novel densovirus from the mosquito Culex pipiens pallens. Journal of General Virology, 89(1), 195–199. https://doi.org/10.1099/vir.0.83221-0

Zhang, Y. Z., Wu, W. C., Shi, M., & Holmes, E. C. (2018). The diversity, evolution and origins of vertebrate RNA viruses. Current Opinion in Virology, 31, 9–16. https://doi.org/10.1016/j.coviro.2018.07.017

Zhang, X., Xie, X., Xia, H., Zou, J., Huang, L., Popov, V. L., Shi, P. Y. (2019). Zika virus NS2A-mediated virion assembly. MBio, 10(5), 1–21. https://doi.org/10.1128/mBio.02375-19

Zirkel, F., Kurth, A., Quan, P.-L., Briese, T., Ellerbrok, H., Pauli, G., Junglen, S. (2011). An Insect Nidovirus Emerging from a Primary Tropical Rainforest. MBio, 2(3), 21–23. https://doi.org/10.1128/mBio.00077-11

Zirkel, F., Roth, H., Kurth, A., Drosten, C., & Ziebuhr, J. (2013). Identification and Characterization of Genetically Divergent Members of the Newly Established Family Mesoniviridae. 87(11), 6346–6358. https://doi.org/10.1128/JVI.00416-13

Chapter 2. A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal[1]

Published as:

Silva, M., Morais, P., Maia, C., de Sousa, C. B., de Almeida, A. P. G., & Parreira, R. (2019). A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal. Virus Research, 274 (July), 197769. https://doi.org/10.1016/j.virusres.2019.197769

[1] This paper was published before a recent taxonomy revision of the *Parvoviridae* family by Penzes et al. in 2020, where all brevidensovirus were renamed as brevihamaparvovirus. In all other sections of this thesis, as well as in most recent papers, the new term is used.

A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal

Abstract

This work describes the detection and partial characterization of mosquito-borne virus genomic sequences, based on the analysis of mosquitoes collected from the Spring to Fall of 2018 in the Algarve (southern Portugal). The viral survey that was carried out using multiple primer sets disclosed the presence of both RNA and DNA viral sequences in these mosquitoes, which were subsequently analyzed using maximum likelihood and Bayesian phylogenetic reconstruction methods. The obtained results brought to light three lineages of insect-specific flaviviruses, a monophyletic cluster of bunyaviruses from an unassigned group within the *Phenuiviridae* family, as well as brevidensoviruses (*Parvoviridae, Densovirinae*). The latter two groups of viruses were here described for the first time in mosquitoes from Portugal. Results relating to the tentative isolation of the putative viruses identified in C6/36 cells are also shown, and the serendipitous, although not unexpected, isolation a Negev-like Nelorpivirus from *Culex laticinctcus* mosquitoes is reported.

Keywords: Insect viruses; Mosquitoes; Phylogenetic analysis; Portugal

Short communication

Among invertebrates, mosquitoes are frequently the focus of viral surveys because they may serve as vectors for many pathogenic agents with (re)emerging potential, including viruses (Gould et al., 2017). Despite their potential to transmit viral agents which may affect human health, mosquitoes have also been shown to harbor many others that seem to display restricted replication capacity in vertebrate cells. These viruses are regarded as insect-specific (Calisher and Higgs, 2018; Junglen and Drosten, 2013), are genetically diverse, and have been tentatively placed in a multitude of viral *taxa* (Abudurexiti et al., 2019; Bolling et al., 2015).

This report describes the results of a survey that was carried out aiming at the detection of a selection of both RNA and DNA viruses, including flaviviruses, phleboviruses, and densoviruses. We based our analysis on mosquitoes recently collected in the Algarve, the southernmost region of the country. This region is climatically influenced by its proximity to the Mediterranean Sea, is a hotspot for tourism, and a temporary haven for migratory birds as they fly to/from Africa/northern Europe. Furthermore, the Algarve displays a combination of ecological and climatic conditions that support the development of multiple species of mosquitoes to high densities, some of which may serve as vectors for arboviruses (Almeida et al., 2008).

The mosquitoes that were analyzed in this work were collected between April and November of 2018 in the district of Faro and corresponded to a convenience sample obtained using CDC-light traps that were not baited with CO2. Due to logistic constraints, the collected mosquitoes were maintained at −20 °C until they were brought to IHMT in Lisbon, where their morphological identification was carried out on ice-bricks using appropriate identification keys (Becker et al., 2010; Ribeiro and Ramos, 1999). These mosquitoes were grouped according to species, sex, geographic origin, and blood-feeding status, and divided into pools with a minimum of 5, and a maximum of 60 specimens. The detection of viral genomes was carried out exclusively using female mosquitoes.

The preparation of mosquito homogenates, nucleic acids extraction, cDNA synthesis, PCR amplification, and DNA cloning was performed as previously described (Carapeta et al., 2015; Pimentel et al., 2019). Detection of *Flavivirus ns5* sequences was carried out using previously described primers and reaction conditions (Vázquez et al., 2012).

Bunyavirus L-coding sequences were targeted for amplification using *Phlebovirus* and *Orthobunyavirus* primers and reaction conditions either previously described (Matsuno et al., 2015; Pereira et al., 2017), or defined in the course of this work. Densovirus sequences were amplified using primers targeting the viral NS1 encoding gene. All the primers, as well as the thermal profiles used for PCR, are listed in Supplementary Table 1. Virus isolation in cell culture was carried out using the *Aedes albopictus* C6/36 cell line, as described before (Carapeta et al., 2015).

Multiple alignments of nucleotide (nt) or amino acid (aa) sequences were performed using the iterative G-INS-I method, as implemented in MAFFT vs. 7. Also, phylogenetic analyses using the maximum likelihood optimization criterium or following a Bayesian approach were carried out essentially as described in previous reports (Pereira et al., 2017; Pimentel et al., 2019).

For molecular confirmation of the morphological identification of mosquitoes, partial mitochondrial cytochrome c oxidase subunit I (COI) sequences were obtained and analysed as previously described (Parreira et al., 2012). However, this analysis was only performed on the pools of mosquitoes where molecular screenings suggested the presence of a viral genome (see below). In all cases, it confirmed the morphological identifications that had been performed. All the nt sequences obtained in the course of this study were deposited in the public sequence databases (GenBank/ENA/DDBJ consortium) under accession numbers LC480777-LC480779 (*ns5*-flaviviruses), LC480766-LC480776 (*L*-bunyaviruses), LC483875 (*ORF1*-Negev-like virus), LC486533 and LC486534 (*NS1*-brevidensoviruses), and LC480766-LC480779, and LC484858 (COI).

The mosquitoes analyzed in this work totaled 2837 specimens (Supplementary Table 2). Most were female (80%, 2276/2837), and the majority were unfed, with only 14.9% (340/2276) evidencing a bloodmeal. They were classified into 6 genera (*Anopheles*, *Culex*, *Culiseta*, *Aedes*, *Coquillettidia*, and *Uranotaenia*), and 16 species. The genus *Culex* encompassed the largest number of specimens distributed into 5 species [*Culex pipiens* s. l. (Linnaeus, 1758), *Cx. theileri* (Theobald, 1903), *Cx. laticinctus* (Edwards, 1913), *Cx. univittatus* (Theobald, 1901), *Cx. hortensis* (Ficalbi, 1889)].

Among the 2276 female mosquitoes that were collected, 79.1% (n=1801) were associated into 50 pools, all of which were subsequently processed for viral screening using a

combination of different PCR/RT-PCR protocols (Supplementary Table 1). Flavivirus and Phlebovirus-like genomes were detected in multiple pools of *Aedes, Anopheles, Culiseta,* and *Culex* mosquitoes, as described in Supplementary Table 2. While the presence of *Orthobunyavirus* sequences could not be unambiguously confirmed in any of the pools analysed, the use of *Brevidensovirus*-specific primers allowed the observation of the expected amplification products when cDNA extracts prepared from *Cx. laticinctus* (n=1) and *Cs. longiareolata* (Macquart, 1838) (n=1) macerates were used.

Flavivirus-specific amplicons were obtained from four species of mosquitoes indicating the presence of Flavivirus genomes in *Cx. laticinctus*, *Cs. annulata* (Schrank, 1776), *Ae. caspius*, and *An. petragnani* (Del Vecchio, 1939). However, attempts to obtain a high-quality sequence from *An. petragnani* (Del Vecchio, 1939) systematically failed, even when recombinant plasmid clones carrying the *Flavivirus*-specific amplicon were used as template for cycle-sequencing, probably due to very low plasmid copy number. On the other hand, analysis of the obtained sequence data (Supplementary Fig. 1) clearly showed that they clustered among the so-called classical insect-specific flaviviruses (cISF), but segregated in three genetically distinct lineages. Two viral sequences detected in pools of *Cx. laticinctus* and *Ae. caspius* were associated with previously identified genetic clusters of viruses circulating in the Iberian Peninsula (Ferreira et al., 2013; Parreira et al., 2012; Vázquez et al., 2012). In addition, a viral sequence obtained from *Cs. annulata* showed high identity (> 98% by BLASTn), and shared common ancestry with another one (KU958176) recently obtained from *Cs. annulata* mosquitoes from Turkey (Ergünay et al., 2017). Curiously, high sequence identity also extended to two other viral sequences (> 96% identity with JF707859-JF707860 using BLASTn) previously found to be integrated within the genomes of *Ae. vexans* (Meigen, 1830) mosquitoes from Spain (Vázquez et al., 2012).

Amplicons with a size compatible with the presence of a *Phlebovirus* L segment were detected in some of the pools analyzed corresponding to five species of mosquitoes (Supplementary Table 2). Surprisingly, high sequence-identity *Phlebovirus* homologs could not be found in the databases when the obtained sequences were analysed with BLASTn/x. This suggested they had been amplified from *Phlebovirus*-like viruses, but not legitimate phleboviruses. Furthermore, phylogenetic reconstruction using nt alignments placed all of them (n=11) outside the *Phlebovirus*, *Banyangvirus*, *Bandavirus*,

and *Goukovirus* genera (not shown). Since these sequences diverged from any of the taxa mentioned above, their identity was investigated using phylogenetic analysis performed on aligned datasets of amino acid sequences of the viral-encoded RNA polymerases (L protein) from the viral groups that compose the Order *Bunyavirales*. Regardless of the method/parameters used for phylogenetic reconstruction, the obtained trees displayed congruent topologies that placed the viral sequences obtained in this study within the *Phenuiviridae* family (Fig. 1). Within this radiation of bunyabiruses, they formed a strongly supported monophyletic cluster that also included viral sequences previously detected using metagenomics/NGS (Chandler et al., 2015; Li et al., 2015; Sadeghi et al., 2018). This viral lineage remains unnamed, as it has not been yet assigned any official designation by the International Committee on Taxonomy of Viruses. Although some of the reference sequences within this cluster had been previously appointed as members of the *Peribunyaviridae* family (Sadeghi et al., 2018), the analysis shown here contradicts that statement.

The analysis of the obtained densovirus NS1 sequences placed them within the *Brevidensovirus* genus (Fig. 2A), while the analysis of a *Brevidensovirus*-only nt sequence dataset (Fig. 2B) revealed that the sequences here described from *Cx. laticinctus* and *Cs. longiareolata* shared a common ancestor with those from brevidensoviruses previously identified in mosquitoes from Russia and Brazil (accession numbers M37899 and GU452799, respectively).

Seven macerates from six species of mosquitoes (*Ae. berlandi, Ae. caspius, An. petragnani, Cs annulata, Cs. longiareolata,* and *Cx. laticinctus*) were selected for viral isolation in C6/36 cells. After two weeks of culture, and when compared with the negative controls, C6/36 exposed to a *Cx. laticinctus* macerate revealed evident CPE. This was characterized by cell growth arrest, cell rounding and detachment from the solid surface (Supplementary Fig. 2A). Somewhat surprizingly, when screened by PCR/RT-PCR using the same primers employed for viral genome screening, none of the culture supernatants revealed the presence of any of the targeted virus-groups. However, the observed CPE recalled previous virus isolation attempts carried out in our laboratory, and suggested the presence of a nelorpivirus in the culture supernatant. This was confirmed using Negev-like virus-specific primers combined with a phylogenetic analysis of the obtained partial ORF1 sequence (Supplementary Fig. 2C). The low success rate of isolation of viruses

using C6/36 cells must take into account the fact that only one blind passage was performed. Although this strategy may have conditioned the possibility of obtaining high titer viral suspensions, the success of viral isolation may have been more seriously compromised by the fact that the mosquitoes were maintained at −20 °C from the day of their collection up to the point when they were identified and macerated. While this does not seem to have affected the infectivity of Negev-like viruses, it may have influenced that of the other viruses detected in the mosquitoes that were analyzed.

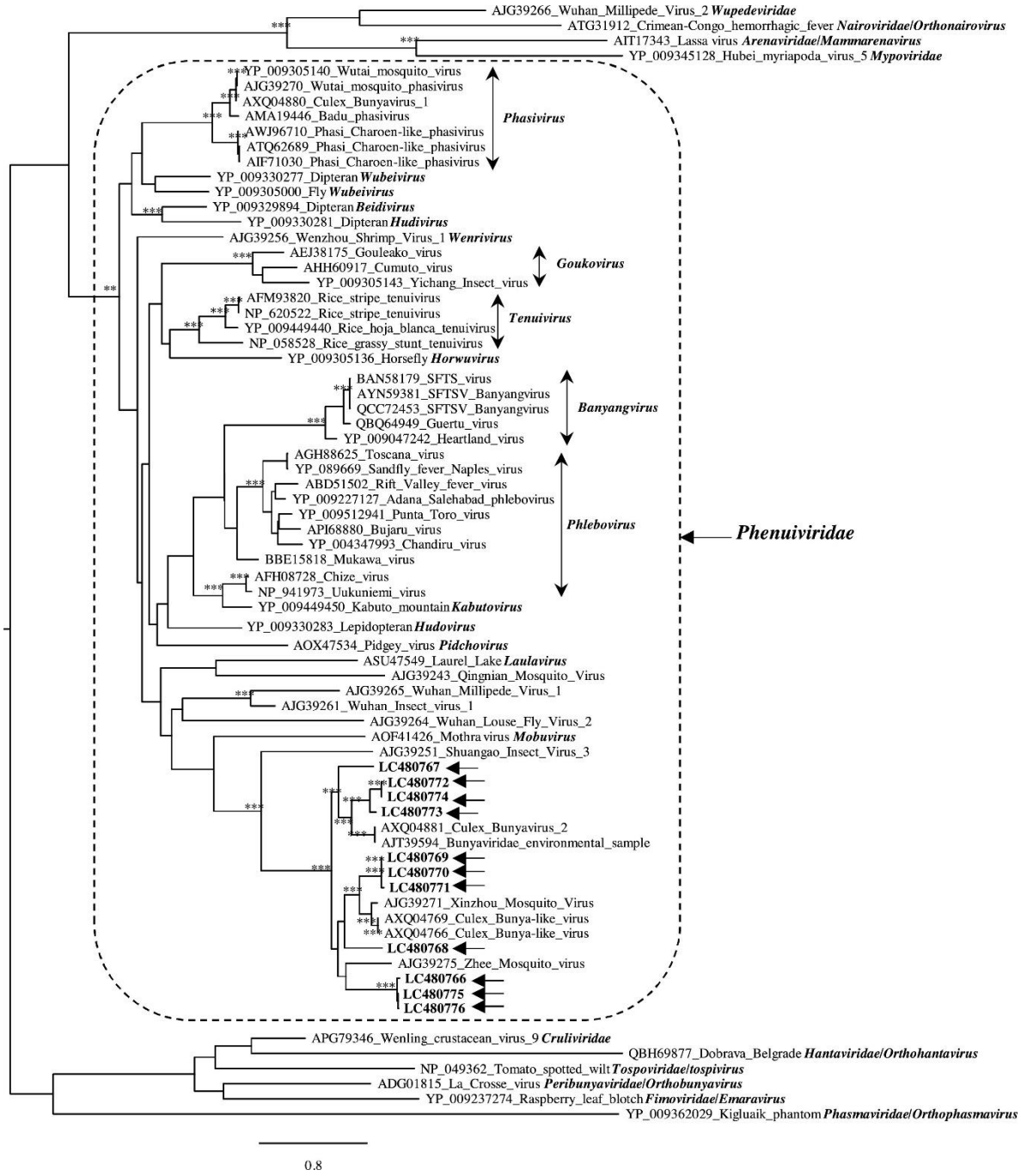


Fig. 1: Phylogenetic analysis of partial amino acid sequences of the viral-ended RNA polymerase of viruses within the Order *Bunyavirales*. At specific branches the number of "*" indicates the support revealed by the different phylogenetic reconstructions methods used, assuming as relevant bootstrap values ≥75% and posterior probability values ≥0.80. Some of the taxonomic groups (viral families and genera) are indicated in boldface and by vertical arrows. The sequences are indicated by their accession number virus name. The sequences described in this work are indicated by their accession numbers, highlighted in boldface, and signaled by the horizontal arrows. The size bar indicates the number of amino acid substitutions per site.

To conclude, this report brought to evidence the circulation of a diverse array of viruses in mosquitoes collected in southern Portugal. While bona fide arboviruses were not identified, three lineages of cISF were described in as many different species of *Culex, Aedes* and *Culiseta* mosquitoes. Two of these lineages had been described previously (Ferreira et al., 2013; Parreira et al., 2012; Vázquez et al., 2012), but another one associated with *Culiseta* specimens had not been described before in the Iberian Peninsula. In addition, the presence of genus-unassigned phenuiviruses (*Buniavirales*) and brevidensoviruses (*Parvoviridae, Densovirinae, Brevidensovirus*) were here described for the first time. Some, or even all, of these viruses, may correspond to viral mutualistic symbionts that are part of the mosquito microbiota, as previously described (Roossinck, 2011).

Author contributions

MS was involved in mosquito identification and processing, as well as molecular screening and phylogenetic analysis; PM was responsible for screening of densoviruses and the analysis of their sequences; CM and CBS were involved in mosquito collection; APGA supervised mosquito identification and processing and manuscript writing; RP supervised all the laboratory tasks involved in this work, including phylogenetic analysis, and wrote the manuscript.

Research involving human and animal participants

This article did not contain any study with human participants or animals.

Declaration of Competing Interest

All authors declare that they have no conflict of interest.

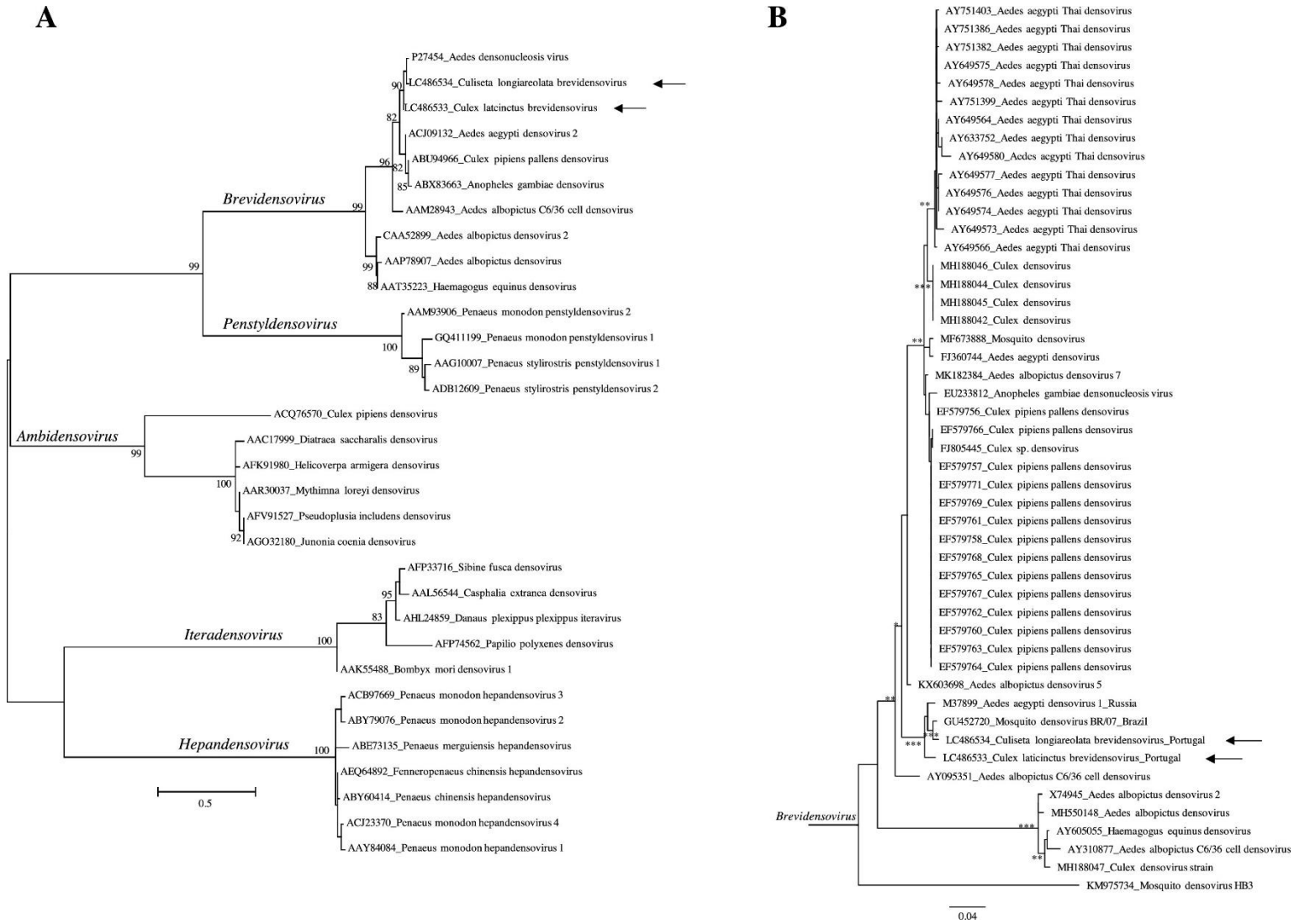


Fig. 2: Phylogenetic analysis of partial NS1 amino acid (A) or nucleotide (B) sequences of viruses in the sub-family Densovirinae (family *Parvoviridae*)(A) or the Brevidensovirus genus (B). In (A) only a maximum likelihood tree is shown, with bootstrap values (≥75%) indicated at specific branches. In (B), at specific branches the number of "*" indicates the support revealed by the different phylogenetic reconstructions methods used, assuming as relevant bootstrap values ≥75% and posterior probability values≥0.80. The sequences obtained in this work are highlighted in boldface and indicated by the horizontal arrows. The size bar indicates the number of amino acid (A) or nucleotide (B) substitutions per site.

Acknowledgments

CM has the support of the Portuguese Ministry of Education and Science (via FCT-Fundação para a Ciência e a Tecnologia, I.P.), through an Investigator Starting Grant (IF/01302/2015). Global Health and Tropical Medicine Center is funded through FCT contract UID/Multi/04413/2013. We would also like to acknowledge the collaboration of a large number of people who have silently participated in this study, by allowing the placement of CDC-light traps in the properties they owned or managed.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.virusres.2019.197769.

References

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S.V., Avšič-Županc, T., Ballinger, M.J., et al., 2019. Taxonomy of the order Bunyavirales: update 2019. Arch. Virol. 164, 1949–1965. https://doi.org/10.1007/s00705-019-04253-6.

Almeida, A.P.G., Galão, R.P., Sousa, C.A., Novo, M.T., Parreira, R., Pinto, J., Piedade, J., Esteves, A., 2008. Potential mosquito vectors of arboviruses in Portugal: species, distribution, abundance and West Nile infection. Trans. R. Soc. Trop. Med. Hyg. 102, 823–832. https://doi.org/10.1016/j.trstmh.2008.03.011.

Becker, N., Petric, D., Zgomba, M., Boase, C., Madon, M., Dahl, C., Kaiser, A., 2010. Mosquitoes and Their Control. Springer, London. Bolling, B.G., Weaver, S.C., Tesh, R.B., Vasilakis, N., 2015. Insect-specific virus discovery: significance for the arbovirus community. Viruses 7, 4911–4928. https://doi.org/10.3390/v7092851.

Calisher, C.H., Higgs, S., 2018. The discovery of arthropod-specific viruses in hematophagous arthropods: an open door to understanding the mechanisms of arbovirus and

arthropod evolution? Annu. Rev. Entomol. 63, 87–113. https://doi.org/10.1146/annurev-ento-020117.

Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., de Matos, A.P., Piedade, J., de Almeida, A.P.G., Parreira, R., 2015. Negeviruses found in multiple species of mosquitoes from southern Portugal: isolation, genetic diversity, and replication in insect cell culture. Virology 483, 318–328. https://doi.org/10.1016/j.virol.2015.04.021.

Chandler, J.A., Liu, R.M., Bennett, S.N., 2015. RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. Front. Microbiol. 6, 1–16. https://doi.org/10.3389/fmicb.2015.00185.

Ergünay, K., Litzba, N., Brinkmann, A., Günay, F., Sarıkaya, Y., Kar, S., Örsten, S., Öter, K., Domingo, C., Erisoz Kasap, Ö., Özkul, A., Mitchell, L., Nitsche, A., Alten, B., Linton, Y.M., 2017. Co-circulation of West Nile virus and distinct insect-specific flaviviruses in Turkey. Parasit. Vectors 10, 1–14. https://doi.org/10.1186/s13071-017-2087-7.

Ferreira, D.D., Cook, S., Lopes, A., de Matos, A.P., Esteves, A., Abecasis, A., de Almeida, A.P.G., Piedade, J., Parreira, R., 2013. Characterization of an insect-specific flavivirus (OCFVPT) co-isolated from Ochlerotatus caspius collected in southern Portugal along with a putative new Negev-like virus. Virus Genes 47, 532–545. https://doi.org/10.1007/s11262-013-0960-9.

Gould, E., Pettersson, J., Higgs, S., Charrel, R., de Lamballerie, X., 2017. Emerging arboviruses: Why today? One Health 4, 1–13. https://doi.org/10.1016/j.onehlt.2017.06.001.

Junglen, S., Drosten, C., 2013. Virus discovery and recent insights into virus diversity in arthropods. Curr. Opin. Microbiol. 16, 507–513. https://doi.org/10.1016/j.mib.2013.06.005.

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., Zhang, Y.Z., 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife 2015, 1–26. https://doi.org/10.7554/eLife.05378.

Matsuno, K., Weisend, C., Kajihara, M., Matysiak, C., Williamson, B.N., Simuunza, M., Mweene, A.S., Takada, A., Tesh, R.B., Ebihara, H., 2015. Comprehensive

molecular detection of tick-borne phleboviruses leads to the retrospective identification of taxonomically unassigned bunyaviruses and the discovery of a novel member of the genus Phlebovirus. J. Virol. 89, 594–604. https://doi.org/10.1128/JVI.02704-14.

Parreira, R., Cook, S., Lopes, Â., de Matos, A.P., de Almeida, A.P.G., Piedade, J., Esteves, A., 2012. Genetic characterization of an insect-specific flavivirus isolated from Culex theileri mosquitoes collected in southern Portugal. Virus Res. 167, 152–161. https://doi.org/10.1016/j.virusres.2012.04.010.

Pereira, A., Figueira, L., Nunes, M., Esteves, A., Cotão, A.J., Vieira, M.L., Maia, C., Campino, L., Parreira, R., 2017. Multiple Phlebovirus (Bunyaviridae) genetic groups detected in Rhipicephalus, Hyalomma and Dermacentor ticks from southern Portugal. Ticks Tick Borne Dis. 8, 45–52. https://doi.org/10.1016/j.ttbdis.2016.09.015.

Pimentel, V., Afonso, R., Nunes, M., Vieira, M.L., Bravo-Barriga, D., Frontera, E., Martinez, M., Pereira, A., Maia, C., Paiva-Cardoso, M., Das, N., Freitas, F.B., Abecasis, A.B., Parreira, R., 2019. Geographic dispersal and genetic diversity of tick-borne phleboviruses (Phenuiviridae, Phlebovirus) as revealed by the analysis of L segment sequences. Ticks Tick Borne Dis. 10, 942–948. https://doi.org/10.1016/j.ttbdis.2019.05.001.

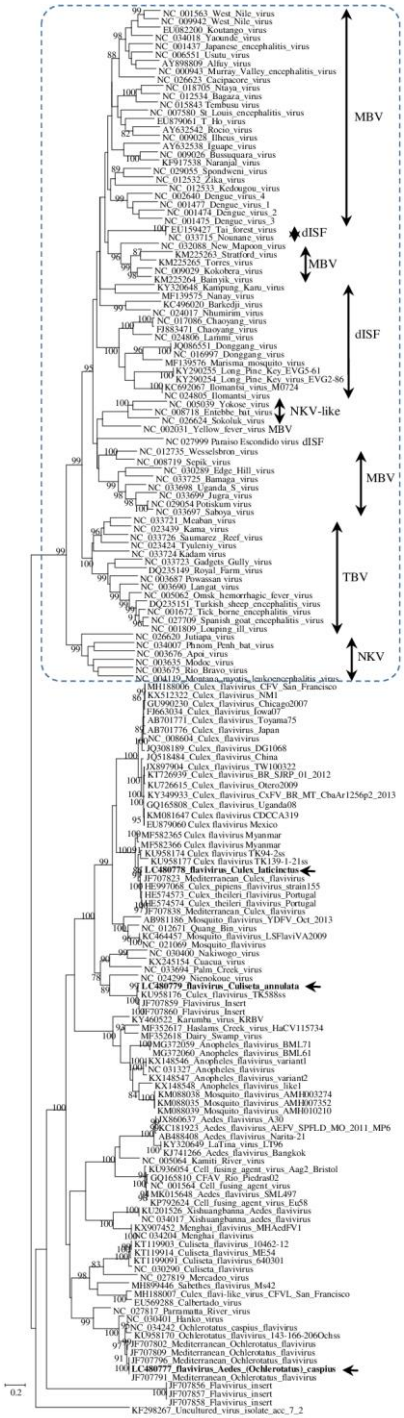
Ribeiro, H., Ramos, H.C., 1999. Identification keys of the mosquitoes (Diptera: culicidae) of continental Portugal, Açores and Madeira. Eur. Mosq. Bull. 3, 1–11. Roossinck, M.J., 2011. The good viruses: viral mutualistic symbioses. Nat. Rev. Microbiol. 9, 99–108. https://doi.org/10.1038/nrmicro2491.

Sadeghi, M., Altan, E., Deng, X., Barker, C.M., Fang, Y., Coffey, L.L., Delwart, E., 2018. Virome of & 12 thousand Culex mosquitoes from throughout California. Virology 523, 74–88. https://doi.org/10.1016/j.virol.2018.07.029.

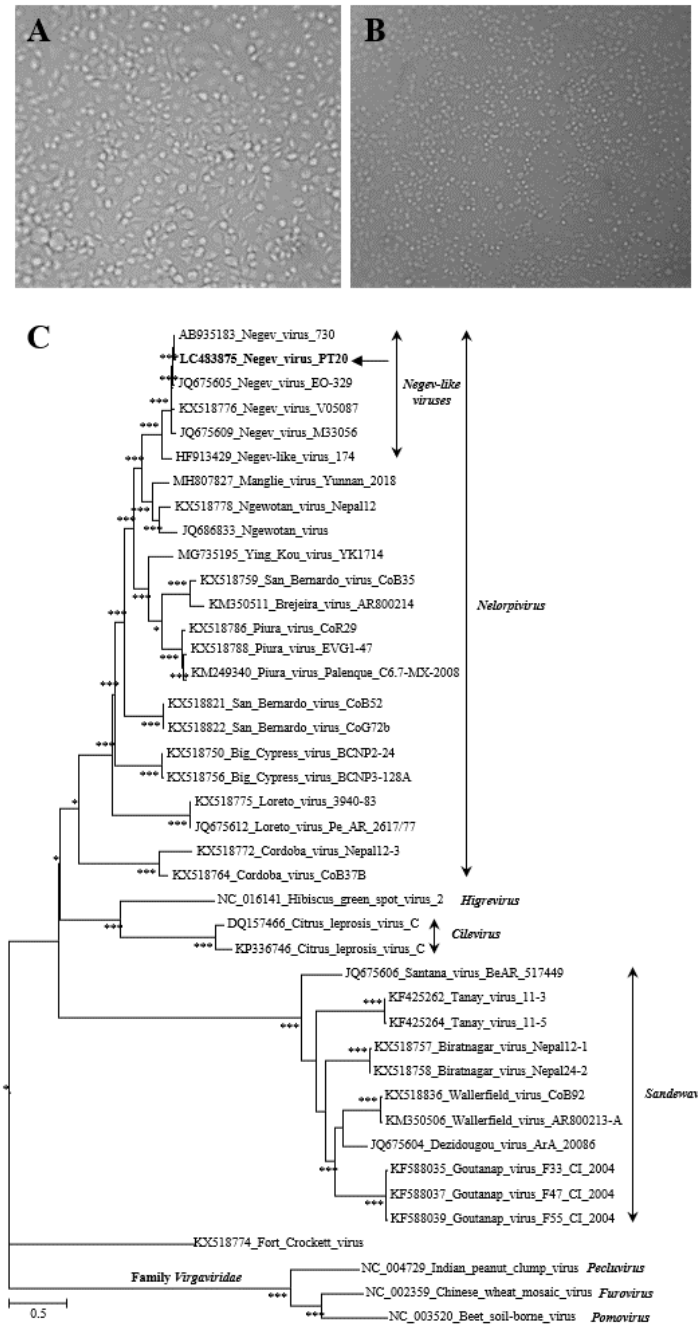
Vázquez, A., Sánchez-Seco, M.-P., Palacios, G., Molero, F., Reyes, N., Ruiz, S., Aranda, C., Marqués, E., Escosa, R., Moreno, J., Figuerola, J., Tenorio, A., 2012. Novel flaviviruses detected in different species of mosquitoes in Spain. Vector-Borne Zoonotic Dis. 12, 223–229. https://doi.org/10.1089/vbz.2011.0687.

A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal

Supplementary Fig. 1: Maximum likelihood phylogenetic analysis of partial *Flavivirus ns5* nucleotide sequences. At specific branches, relevant bootstrap values (≥75%) are indicated. The multiple reference sequences used include mosquito-borne viruses (MBV), tick-borne viruses (TBV), no known vector viruses (NKV), dual-host associated insect-specific viruses (dISF), and classical insect-specific flaviviruses (cISF). The sequences described in this work are indicated in boldface and the horizontal arrow. All the sequences used are designated by their respective accession numbers_virus name_strain (when available). The size bar indicates the number of nucleotide substitutions per site.



Supplementary Fig. 2: Microscopic observation of C6/36 cells mock-infected cells (A; 300×), or after infection (day 2) with a Negev-like virus isolated from *Cx. laticinctus* (B; 300×). Phylogenetic analysis of partial *ORF1* nucleotide sequences of viruses from the family *Virgaviridae*, the genera *Higrevirus* and *Cilevirus*, and the proposed genera *Sandewavirus* and *Nelorpivirus*. In the latter, the group formed by Negev-like viral sequences, and that included the sequence of the virus isolated in the course of this work (boldface and signaled by the horizontal arrow), is also indicated. At specific branches the number of "*" indicates the support revealed by the different phylogenetic reconstructions methods used, assuming as relevant bootstrap values ≥75% and posterior probability values ≥0.80. The sequences used are indicated by their accession number_virus name. The size bar indicates the number of nucleotide substitutions per site.



Supplementary Table 1: PCR primers and thermal profiles used in this work.

Target gene	Primer sequence (5'-3')	Thermocycling conditions	Reference
ns5 (Flaviviruses)	1st PCR 1NS5F: GCATCTAYAWCAYNATGGG 1NS5R: CCANACNYNRTTCCANAC	95 ºC - 5 min; 45 cycles [95 ºC - 1 min; 50 ºC - 4 min; 72 ºC - 1 min]; 72 ºC - 5 min	Vásquez et al., 2012
	2nd PCR 2NS5F: GCNATNTGGTWYATGTGG 2NS5R: CATRTCTTCNGTCGTCATCC	Same as above	
L (Phleboviruses)	TBPVL2759F: CAGCATGGIGGICTIAGAGAGAT TBPVL3267R: TGIAGIATSCCYTGCATCAT HRT-GL2759F: CAGCATGGIGGIYTIAGRGAAATYTATGT HRT-GL3276R: GAWGTRWARTGCAGGATICCYTGCATCAT	95 ºC - 2 min; 45 cycles [95 ºC - 30 sec; 55 ºC - 30 sec; 72 ºC - 30 sec]; 72 ºC - 5 min	Matsuno et al., 2015
L (Orthobunyaviruses)	1st PCR OrthoF12: TRACTGARCCWTCTMGATATATGATAATGAAYT OrthoR1: CATCTTGDGCACTCCATTTTGACATRTCHGC	95 ºC - 2 min; 45 cycles [95 ºC - 30 sec; 53 ºC - 45 sec; 72 ºC – 45 sec]; 72 ºC - 5 min	This work
	2nd PCR OrthoF12: TRACTGARCCWTCTMGATATATGATAATGAAYT OrthoR2: CATACATRCACATYTTDGCTTCAAATTC	Same as ab	
NS1 (Densovirus)	1st PCR DF1: AACCGTTGGTGACCTCTACCCACAATTAC DR1: CGGATGCAATAGAGAATGAAGTTCCTGAG	95 ºC - 2 min; 45 cycles [95 ºC – 30sec; 53 ºC – 30 sec; 72 ºC – 45 sec]; 72 ºC - 5 min	This work
	2nd PCR DF2: GAAACTATGGATAATAACGGGTCACAGG DR2: CGCTTCTGCACTTCCTGCGCTTGTCGC	Same as ab	
ORF1 (Negev-like viruses)	1st PCR NegeF: CAYGTRAARATYTTCTGCGAYATGTC NegevinR: TAATCGTTTGTGCGGTARACATTGAGGC	95 ºC - 2 min; 45 cycles [95 ºC - 30 sec; 55 ºC - 30 sec; 72 ºC – 1 min]; 72 ºC - 5 min	Carapeta et al., 2015
	2nd PCR NegevinF: AGTGCTTCAACGTGACATTCCCCCGTCC NegevinR: TAATCGTTTGTGCGGTARACATTGAGGC	Same as above	
ORF1 (Loreto-like viruses)	1st PCR LorF: CGGCAATTTGGAATCGAAGAGGAACTTGTC LorRout: CCACATGAAGGAGGAAGTGTACAACC	95 ºC - 2 min; 45 cycles [95 ºC - 30 sec; 55 ºC - 30 sec; 72 ºC – 1 min]; 72 ºC - 5 min	Carapeta et al., 2015
	2nd PCR LorF: CGGCAATTTGGAATCGAAGAGGAACTTGTC LorR: TGTGCGATGAACTTCGATACATTCCGGGTC	Same as above	
ORF1 (Denzidougou-like viruses)	1st PCR DenzF: TAATTTGTGYGTTACYGCTCTKACTMGGCACAC DenzR: ATACGAACYTTRGGATTRCGTTTCAGAGAC	95 ºC - 2 min; 45 cycles [95 ºC - 30 sec; 55 ºC - 30 sec; 72 ºC – 1 min]; 72 ºC - 5 min	Carapeta et al., 2015
	2nd PCR DenzF: TAATTTGTGYGTTACYGCTCTKACTMGGCACAC DenzinR: GCKGGAGCAGGAGTGCTCAACMMCGG	Same as above	
COI	LCO1490: GGTCAACAAATCATAAAGATATTGG HC02198:TAAACTTCAGGGTGACCAAAAAATCA	95 ºC - 5 min; 40 cycles [95 ºC - 30 sec; 48 ºC – 30 sec;	Folmer et al., 1994,

	72 ºC – 45 sec]; 72 ºC - 5 min	Cook et al., 2009

I-inosine, M-A or C, N- any base, R-A or G, S-C or G, W-A or T, Y-C or T. Sequences in bold-face are complementary to the genome of bacteriophage SPP1 (X97918).

Supplementary Table 2: Mosquito collections and viral detection analyses.

Mosquito species	Total	Mosquitoes analysed (number of pools)	Pools (lab codes)	Flavivirus detection	Phenuivirus detection	Brevidensovirus detectioon
Ae. berlandi	20	13 (n=1)	37[2]	0	1	0
Ae. caspius	401	344 (n=8)	2, 3[1,2], 69, 70[2], 71[1,2], 72[2], 73, 74[2]	2	5	0
Ae. detritus	16	16 (n=1)	4[2]	0	1	0
Ae. echinus	3	0	na	na	na	na
An. atroparvus	860	314 (n=7)	6, 8, 10, 40, 42, 44, 46	0	0	0
An. claviger/petragnani	104	98 (n=3)	48[2], 49[2], 50[1,2]	1	3	0
An. plumbeus	1	0	na	na	na	na
Cq. richiardii	3	0	na	na	na	na0
Cs. annulata	62	55 (n=3)	16[1], 52[1,2], 53[2]	2	2	0
Cs. longiareolata	112	66 (n=3)	18, 54[3], 79	0	0	1
Cx. hortensis	2	0	na	na	na	na
Cx. laticinctus	194	75 (n=3)	19, 20[1,3], 58	1	0	1
Cx. pipiens s.l.	645	474 (n=12)	21, 22, 23, 24, 25, 26, 27, 60, 61, 62, 81, 86	0	0	0
Cx. theileri	282	267 (n=6)	29, 30, 31, 64, 65, 66	0	0	0
Cx. univittatus	162	79 (n=3)	32, 33, 34	0	0	0
Ur. unguiculata	2	0	na	na	na	na
Total	**2869**	**1801 (n=50)**	**na**	**6**	**12**	**2**

Ae: *Aedes*, *An*: *Anopheles*, *Cq*: *Coquillettidia*, *Cs*: *Culiseta*, *Cx*: *Culex*, *Ur*: *Uranotaenia*. [1]Detection of flavivirus sequences; [2]Detection of phenuivirus sequences; [3]Detection of brevidensovirus sequences; na – not applicable; s.l. – sensu lato.

Chapter 3. Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: Genome detection and phylogenetic characterization of viral sequences[2]

Published as:

Morais, P., Pinto, J., Jorge, C. P., Troco, A. D., Fortes, F., Sousa, C. A., & Parreira, R. (2020). Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: Genome detection and phylogenetic characterization of viral sequences. Infection, Genetics and Evolution, 80 (January), 104191. https://doi.org/10.1016/j.meegid.2020.104191

[2] This paper was published before a recent taxonomy revision of the *Parvoviridae* family by Penzes et al. in 2020, where all brevidensovirus were renamed as brevihamaparvovirus. In all other sections of this thesis, as well as in most recent papers, the new term is used.

Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: Genome detection and phylogenetic characterization of viral sequences

Abstract

This report describes a survey of RNA and DNA viruses carried out in adult mosquitoes from Angola, raised under laboratory conditions from field-collected immature forms. This viral genomic survey was performed using different sets of primers targeting groups of arboviruses with a considerable impact on human health, including flaviviruses, alphaviruses, and phleboviruses. Furthermore, the viral survey that was performed also included detection of densoviruses. The obtained results did not reveal the presence of recognizable pathogenic arboviruses but allowed the identification of insect-specific flaviviruses from two genetic lineages and a single lineage of brevidensoviruses. These viruses, collectively detected in *Anopheles* sp. and *Culex pipiens* s.l. mosquitoes, were most probably transmitted vertically.

Keywords: Flaviviruses; Densoviruses; Vertical transmission; Mosquitoes; Phylogenetic analysis; Angola

Short communication

The virome of mosquitoes (Diptera: Culicidae) comprises a multitude of pathogenic viruses, as well as others that seem to display limited replication capacity in vertebrate cells (Calisher and Higgs, 2018; Junglen and Drosten, 2013). These so-called insect-specific viruses comprehend a genetically diverse assemblage of both RNA and DNA viral agents and have been classified in a wide variety of distinct taxa (Nouri et al., 2018). While these viruses are regarded as non-pathogenic to vertebrates, their impact on the fitness of their invertebrate hosts is still open to debate.

In the wake of the yellow fever virus (YFV) epidemic that took place in Angola in 2015–2016 (Grobbelaar et al., 2016), while surveying the distribution of *Aedes aegypti* (the most probable vector of YFV during the outbreak), a convenience sample of mosquitoes was obtained which served as a starting point for a viral survey. These mosquitoes were collected between September and November 2016, at different sampling points located in the capital city of Luanda (−8° 50′ 12″ S, 13° 14′ 03″ E), and some of its peripheric municipalities. In addition, mosquitoes were also collected in the cities of Sumbe (−11° 12′ 21″ S, 13° 50′ 37″ E), Benguela (−12° 34′ 34 S, 13° 24′ 19 E) and Huambo (−12° 46′ 33 S, 15° 44′ 21 E), in the provinces of Cuanza Sul, Benguela, and Huambo, respectively. The geographic distribution and total of mosquitoes collected in each region are indicated in Supplementary Fig. 1. The analyzed mosquitoes were collected as larvae or pupae by sampling with dips and pipettes, or as eggs laid in ovitraps.

The mosquito immature forms were reared up to the imago stage, and immediately after emergence, male and female mosquitoes were separated, and the specimens were grouped into pools with a minimum of 3 and a maximum of 63 specimens per pool. This grouping took into account the geographic sampling region and the mosquito taxa, with groups being defined as “Anophelinae”, “*Aedes aegypti*”, and “other Culicinae”. For the preparation of the different pools of specimens, each mosquito was briefly soaked in ethanol, dried onto filter paper and immediately stored at −20 °C in RNAlater® (Invitrogen, USA).

Viral detection was carried out using nucleic acids that were extracted with NZYol (NZYtech, Portugal), starting from aliquots of centrifugation clarified mosquito macerates, prepared as previously described (Carapeta et al., 2015; Pimentel et al., 2019).

Multiple combinations of primers were used in PCR-based assays aiming at the molecular detection of different groups of viruses. For the detection of viruses with RNA genomes, prior to PCR, total RNA was converted to cDNA using the NZY First-Strand cDNA Synthesis Kit (NZYtech, Portugal). The primers used for viral detection targeted the genomes of bona fide arboviruses, including flaviviruses (Vázquez et al., 2012), alphaviruses (Sánchez-Seco et al., 2001) and phleboviruses (Matsuno et al., 2015). The presence of densoviruses was also investigated using NS1 and VP-specific primers developed in the course of this study. The primers used are indicated in Supplementary Table 1, where the PCR conditions for detection of densoviruses are also indicated (Supplementary Table 1 foot-note). The obtained amplification products were sequenced either directly, or after cloning in pNZY-28, using the NZY-A PCR cloning kit (NZYtech, Portugal).

Sequence similarity searches were carried out using BLASTn/p, while phylogenetic trees were constructed using maximum likelihood and Bayesian approaches, essentially as described by Pimentel et al. (2019), using the Mega 6.0 (Tamura et al., 2013) and BEASTv1.7.5 (Drummond and Rambaut, 2007) software packages, respectively. For Bayesian analyses, two independent Markov chain Monte-Carlo (MCMC) runs were performed under a relaxed lognormal clock, until 1 × 108 states were sampled, 10% of which were later discarded as burn-in. Logistic and Gaussian Markov random field (GMRF) demographic priors were used for phylogenetic reconstructions under a Bayesian framework. The Tracer software (http://beast.bio.ed.ac.uk/tracer) was used to check for chain-convergence and adequate (> 200) effective sample size (ESS). Phylogenetic trees were sampled on every 10,000th MCMC step, finally summarized as a maximum clade credibility tree (MCC), and visualized with the FigTree v1.4.2 software (http://tree.bio.ed.ac.uk/software/figtree/). Evolutionary divergence values over sequence pairs (intra- and inter- groups of sequences) were calculated with the JTT matrix-based model (as implemented in Mega 6) using NS1 and VP amino acid alignments. Intra-genus nucleotide sequence diversity values were calculated with Mega 6 using the TN93 + Γ model.

The sequences obtained in the course of this work were deposited in the public databases (GenBank/ENA/DDBJ) under accession numbers LC485954-LC485964 (*Brevidensovirus* NS1), LC485294-LC485303 (*Brevidensovirus* VP), and LC485965-

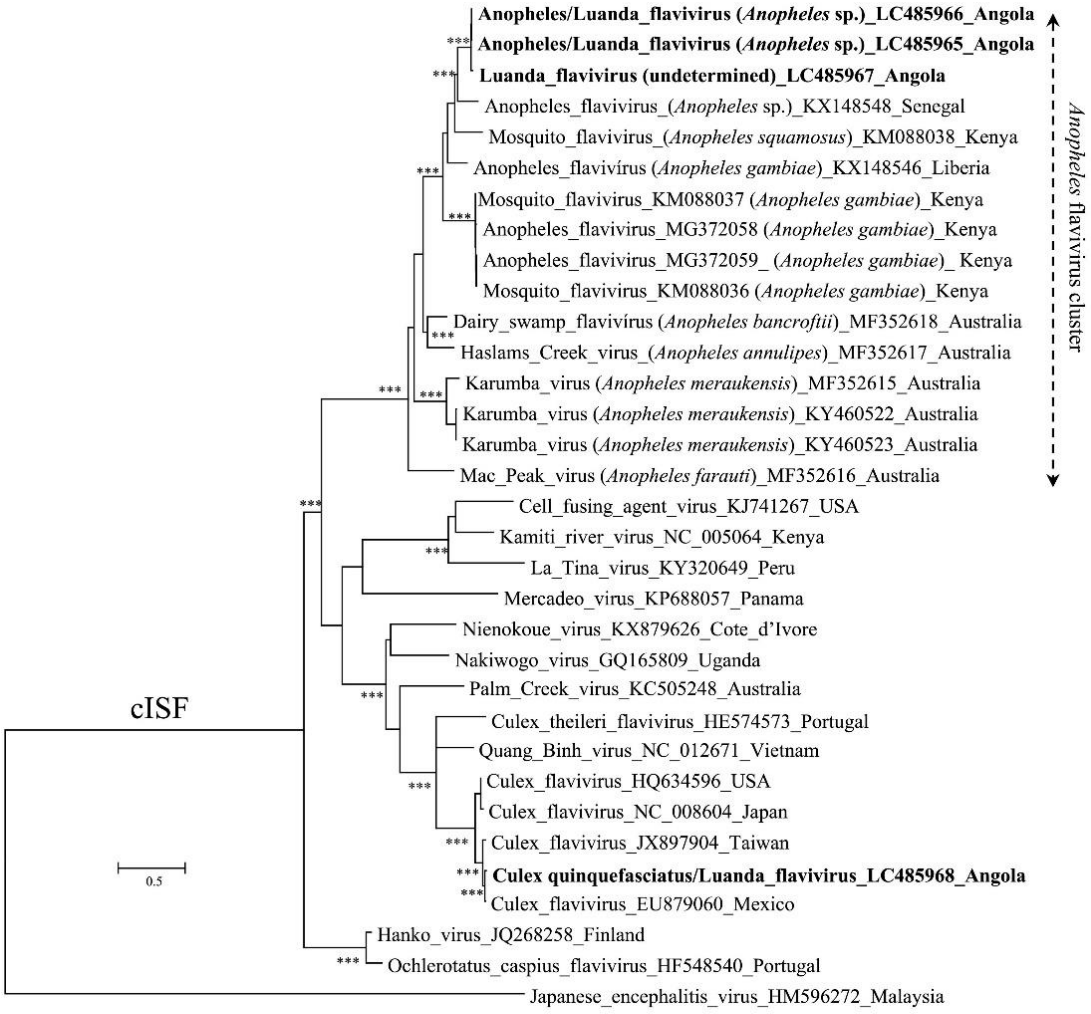
LC485968 (*Flavivirus* NS5). These sequences are listed in Supplementary Table 2a. Sequences LC484845- LC484858 correspond to the mitochondrial cytochrome C subunit I (COI). Others, used as references in the different genetic analyses, were downloaded from GenBank.

Mosquito species identification was carried out for each homogenate from which a viral sequence was obtained (see below). This was achieved by DNA-barcoding using BoldSystems vs4 (Ratnasingham and Hebert, 2007), based on the analysis of partial sequences of the mosquito mitochondrial COI (Cook et al., 2009; Folmer et al., 1994). In addition, a multiplex-PCR targeting the acetylcholinesterase (Ace-2) coding nucleotide sequence was also used to tentatively discriminate between members of the *Culex pipiens* s.l. complex, which includes *Cx. pipiens* (L. 1758, or *Cx. pipiens* sensu stricto), Cx. quinquefasciatus (Say 1823), *Cx. pipiens pallens* (Coquillett 1898), and *Cx. australicus* (Dobrotworsky & Drummond 1953) (Smith and Fonseca, 2004).

In total, 60 mosquito pools were analyzed, corresponding to 1441 mosquitoes. Viral screening using *Alphavirus*- or *Phlebovirus*-specific primers did not reveal the presence of their respective genomes in any of the pools analyzed, as none of the expected virus-specific amplicons were ever observed after gel electrophoresis. However, using cDNA prepared from 5 pools (Supplementary Table 2 a, b), corresponding to both male and female mosquitoes, the presence of a 1 kbp amplicon was observed. This amplicon encompasses part of a flavivirus NS5 coding sequence, that on the Culex flavivirus strain CxFV-Mex07 reference sequence (accession number EU879060) would define a section of the viral genome from coordinates 9800 to 9901. No virus-specific amplicons were ever obtained when reverse-transcription was omitted, or when total DNA was directly used as template for PCR. This seems to dismiss their origin as flavivirus sequences integrated into mosquito genomes (Crochu et al., 2004). Three of these pools corresponded to a morphologically homogeneous group of mosquitoes collected from the same breeding site (Hotel Panorama, Luanda), classified as *Anopheles* sp. on the basis of DNA-barcoding and Blastn searches (Megablast option). Indeed, its closest homolog (sharing 99.82% sequence identity by Blastn analysis) was sequence MF372931. This corresponded to Anopheles sp. 1 YL-201 collected in 2015 in Benguela, suggesting that they correspond to a species of *Anopheles* mosquitoes that seems to be common in Angola. The other two pools where flavivirus sequences were detected were shown to

correspond to either *Culex quinquefasciatus* or a mixture of at least *Cx. pipiens* and *Cx. quinquefasciatus*, as defined by COI-based barcoding, Ace-2 PCR results, and Blast sequence searches. The Ace-2 PCR also revealed a band compatible with the presence of *Cx. pallens*, but given the Asian geographic distribution of this species, their presence in Angola was deemed unlikely.

From five flavivirus-positive pools we obtained four viral sequences (Fig. 1), and their analysis showed that none of them were directly related to pathogenic flaviviruses. On the contrary, the sequences segregated in two distinct genetic lineages of the so-called classical insect-specific flaviviruses, or cISF (Bolling et al., 2015). One of these viral sequences (LC485968), obtained from *Cx. quinquefasciatus* mosquitoes collected in Sumbe, ca. 330 km south of the capital Luanda, was included in the genetic lineage of Culex flaviviruses. The other three were amplified from pools of mosquitoes identified as *Anopheles* sp. or *Culex pipiens* complex mosquitoes (on the basis of COI sequence analysis), all collected in the vicinity of Luanda. These viral sequences were shown to share a common ancestry and formed a distinct phylogenetic lineage, having *Anopheles* flavivirus (KX148548) as their closest homolog (approximately 83% nucleotide identity, see table in Fig. 1).



Accession number	Sequence identity (Blastn best match)	Sequence identity (Blastp best match)
LC485965	83% with Aedes flavivirus (KX148548)	95% Anopheles flavivirus (YP_009305401)
LC485966	83% with Aedes flavivirus (KX148548)	95% Anopheles flavivirus (YP_009305401)
LC485967	83% with Aedes flavivirus (KX148548)	94% Anopheles flavivirus (YP_009305401)
LC485968	98% Culex flavivirus (MH719098.)	100% Culex flavivirus (ACJ64914)

Fig. 1: Phylogenetic analysis of partial Flavivirus NS5 nucleotide sequences. After removing poorly aligned regions from the multiple sequence alignment, the trees were constructed using 837 unambiguously aligned nucleotides. At specific branches, the number of "*" indicates the branch-support as revealed by the different phylogenetic reconstructions methods used, and assuming as relevant bootstrap values ≥75% (using 1000 resampling of the sequence data in maximum likelihood analysis) and posterior probability values ≥0.80 (when Bayesian approaches were used). One, two or three "*" would indicate that a given branch had been supported by one, two, or all the phylogenetic reconstruction approaches used in the analysis (ML and Bayesian analysis using two sets of demographic priors). In the *Anopheles* flavivirus clade (indicated by the lateral, vertical arrow), virus sequence designation includes their origin (mosquito species), when available. The bar indicates the average number of substitutions per site. At the base of the figure, the table indicates, for each NS5 sequence obtained what were their corresponding best matches by sequence similarity searches using Blastn and Blastp. The NS5 sequence from Japanese encephalitis virus from Malaysia (HM596272) was used as the tree outgroup. cISF indicates classical Insect Specific Flaviviruses.

Despite the fact that the molecular identification that was performed indicated *Cx. pipiens* s.l. as the origin for viral sequence LC485967 (one of the three sequences mentioned above), this identification may be misleading. In fact, given the apparent high-species specificity previously suggested for the flaviviruses detected in *Anopheles* mosquitoes (Colmant et al., 2017), the association of an apparently *Culex*-derived viral sequence with this group (Fig. 1) was considered controversial. However, one must bear in mind that Sanger-sequencing, being the starting point for identifications based on sequence analysis of PCR amplicons, is a population-sequencing approach that only reveals the sequence of the most abundant molecular form in a PCR-product. For this reason, while sequence analyses results may have suggested a *Culex* origin for the viral sequence LC485967, this does not formally exclude the possibility that it may have been derived from one (or a small number of) non-*Culex* mosquitoes (possibly *Anopheles*) originally present in the pool in question. In addition, a very low amount of Flavivirus-specific amplificon was obtained, which is also compatible with their origin being a low number of non-*Culex* derived DNA molecules. Since the lack of *Anopheles* genus-specific primers in our lab precludes a PCR-based confirmation for the presence of *Anopheles* sequences in the DNA extract in question, for the time being, the exact origin of LC485967 viral sequence remains undetermined. A clarification of this point would require, for example, a larger sampling of mosquitoes, and the isolation of the detected flaviviruses in *Aedes*, *Anopheles*, and *Culex* cell lines, in order to check for mosquito specificity. Then again, viral replication in cell culture could not have been performed in the context of our work, as the analyzed mosquitoes had been stored in RNAlater®, which compromises viral infectivity.

Among the viruses that may exploit vertical transmission for their natural maintenance stand those commonly known as densoviruses. Although they are not pathogenic to humans, they have very rarely have been described in African mosquitoes. For these reasons, the presence of the genomes of DNA viruses of the *Densovirinae* subfamily (*Parvoviridae*) was also investigated. These viruses comprise a diverse group of linear single-stranded DNA viral agents that infect arthropods, many of which are pathogenic to the members of seven insect Orders, including the Order Diptera (Martynova et al., 2016). Furthermore, these viruses do not replicate in vertebrate cells and may integrate their genomes into their host's (mosquito) chromosomal DNA. In this regard, they can be

exploited as vehicles for stable expression of heterologous proteins in insect cells, host paratransgenesis and even used for mosquito control (Gu et al., 2011; Ren et al., 2008; Tijssen et al., 2016).

The genomes of densoviruses were detected using primers that were designed on the basis of the identification of conserved regions in multiple sequence alignments [obtained using MAFFT vs7; Katoh and Standley, 2013] of viral genomic reference sequences downloaded from the nucleotide sequence databanks. Two nested-PCR protocols were developed, targeting either the NS1 (regulatory protein) or VP (capsid) coding regions (Supplementary Table 1). The screening for the presence of genomes of densoviruses using primers targeting NS1 sequences revealed positive amplification results for 15 out of 60 pools (25%) of mosquitoes (Supplementary Table 2 b). The majority (14 out of 15) included either male or female mosquitoes that had been collected in the city of Luanda (and its surroundings), corresponding either exclusively to *Culex quinquefasciatus* or mixtures of *Cx. pipiens* and *Cx. quinquefasciatus* mosquitoes. One additional pool of female mosquitoes, collected in the city of Sumbe, also revealed the presence of densovirus genomes.

Two of these pools (including the latter) had also been found positive for the presence of *Flavivirus* genomes. For 10 of these pools of mosquitoes, partial VP-encoding sequences were also obtained using specific primers, while we repeatedly failed to amplify these sequences from one of them. A preliminary phylogenetic analysis (maximum likelihood) was carried out using sequence alignments of the putative NS1 and VP translated products encoded of these sequences (Fig. 2A and B, respectively), along with many others from viruses of the five genera of the *Densovirinae* subfamily (Cotmore et al., 2014). Both phylogenetic trees gave similar topologies, where the viral sequences obtained segregated together within the monophyletic cluster defining the genus *Brevidensovirus*.

Furthermore, the inclusion of the VP Angolan *Brevidensovirus* sequences in a single monophyletic cluster was supported by the identification of sequence group-specific polymorphisms (defined by $p < .001$ in a Chi-squared test using Yate's correction) in the sequence of the capsid protein. These include both conservative (K126R, N253T, S254T, V256M), and non-conservative (Y72L, S135A, K218Q, E234A) amino acid substitutions.

On the basis of the analyses described above, the genomes of the densovirus strains detected do not seem to have resulted from intergenus recombination. However, since intragenus genetic recombination has been previously shown to affect the evolutionary history of these viruses (Martynova et al., 2016), viral genomes were further characterized by combining phylogenetic and recombination analyses, using concatenated alignments of the NS1 and VP sequences. As expected, given the topology of the NS1 and VP phylogenetic trees, the inspection of the concatenated multiple sequence alignment with the RDP4 software (Martin et al., 2015) did not reveal any evidence of intra-genus recombination.

The analysis of genetic distances was also carried out using an alignment of concatenated NS1/VP-coding sequences, corrected using the TN93 + Γ model, as suggested by Mega 6 (Tamura et al., 2013). Estimates of evolutionary divergence over amino acid sequence pairs between groups (genera) are shown in Supplementary Table 3.

The amplification of viral sequences from adult mosquitoes that were collected as immature forms and their occurrence in both male and female specimens supports the possibility that all the viruses detected in this study might exploit vertical (transovarian) transmission for their natural maintenance. Nonetheless, since the presence of viral sequences was not investigated, for example, in eggs, the obtained results are also formally compatible with the possibility of horizontal viral transmission between immature mosquito forms at their breeding sites.

While the circulation of identifiable pathogenic flaviviruses (e.g. dengue, Zika or YFV), alphaviruses (e.g. chikungunya) or phleboviruses (e.g. Rift Valley fever virus) was not verified, one should bear in mind that in addition to the analytic bottleneck imposed by selecting for apparent vertical transmission, most of the mosquitoes analyzed were not those most frequently associated with the transmission of the arboviruses, i.e. *Ae. aegypti* and *Ae. albopictus*. However, during the YFV 2016 epidemic in Angola, a double YFV-Japanese encephalitis virus (JEV) infection was detected in the Cunene province in the south of the

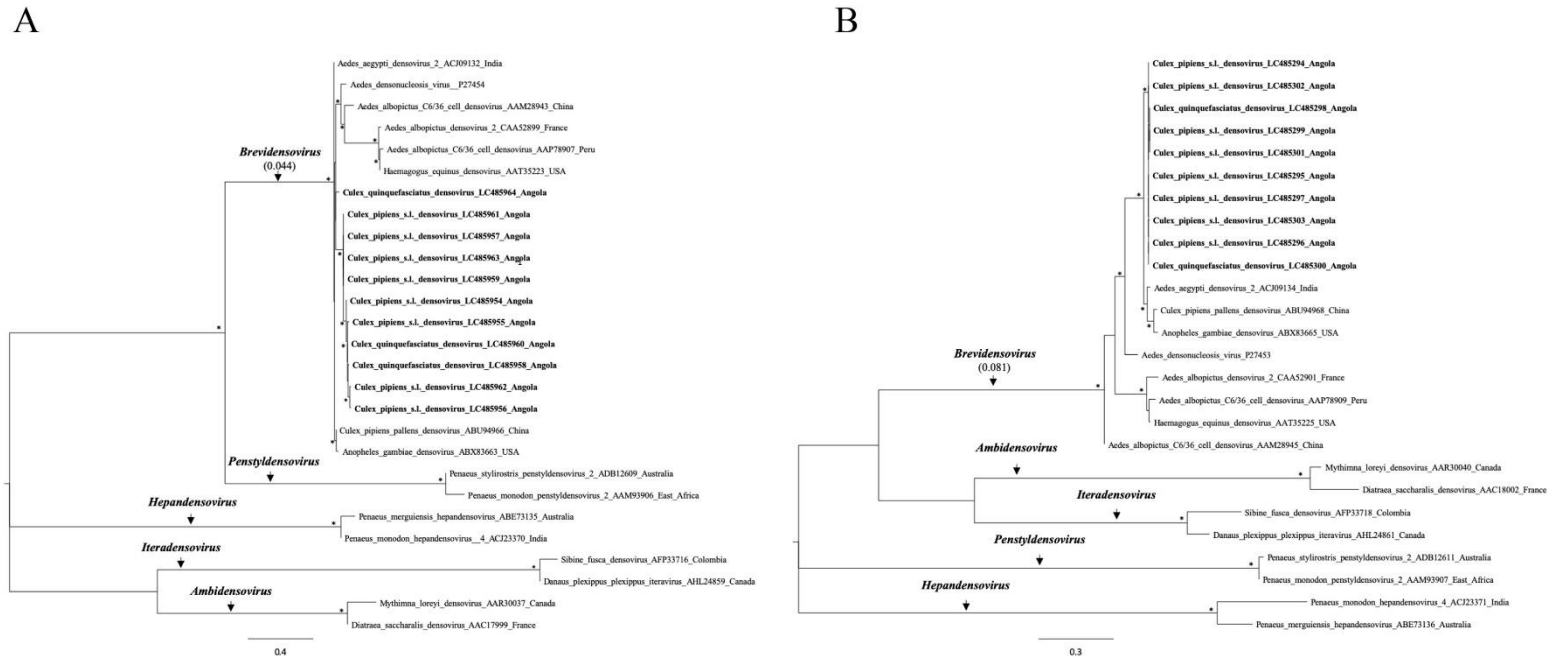


Fig. 2: Phylogenetic analysis of NS1 (A) and VP (B) densovirus amino acid sequences. Trees were reconstructed based on the analysis of approximately 180 (NS1) or 240 (VP) amino acid residues. The branches defining the five genera of the subfamily Densovirinae are indicated. At specific branches, the "*" indicates the support obtained using maximum likelihood, corresponding to bootstrap values ≥75% of an analysis that included 1000 resamplings of the initial sequence data. The values shown below the different viral genus refer to intra-group amino acid sequence divergence (in brackets). The bar indicates the average number of amino acid substitutions per site.

country (Simon-Loriere et al., 2017), suggesting that pathogenic flaviviruses may already co-circulate in Angola. Indeed, this YFV/JEV coinfection was detected by de novo assembly of high throughput sequence data obtained from a single YFV-infected 19-year-old man, with a 5-day history of fever. While this sample was obtained at the Cunene provincial hospital (> 800 km south-east from the capital), at the onset of disease, the patient was working in Luanda and reported no history of traveling abroad. While YFV is usually transmitted by Ae. aegypti, the circulation of JEV is most frequently associated with Culex mosquitoes. Detection of a pathogenic virus in the vector in the absence of sympatrically described cases of symptomatic disease in humans is a rare event. On the contrary, cISF and brevidensoviruses were relatively frequent, altogether being detected in a total of 18 out of 60 pools analyzed (i.e., in 1/3 of the total). To our knowledge, this study corresponds to the first identification of cISF and brevidensoviruses in mosquitoes from Angola. As the specimens analyzed corresponded to a small convenience sample, the obtained results suggest that many other viral strains wait to be identified. While the viruses here described in mosquitoes from Angola may correspond to viral mutualistic symbionts that are part of the mosquito microbiota, as previously described for many other organisms, their presence may interfere with their host's competence for transmission of bona fide arboviruses (Nouri et al., 2018). This subject deserves to be investigated in the future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to acknowledge the silent collaboration of those who have participated in the collection of mosquitoes analyzed in the course of this study. This work received funding from Camões - Instituto da Cooperação e da Língua (Ministry of Foreign Affairs, Portugal), the Ministry of Health of Angola and from Fundação para a Ciência e a Tecnologia (FCT) for funds to GHTM-UID/Multi/04413/2019.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2020.104191.

References

Bolling, B.G., Weaver, S.C., Tesh, R.B., Vasilakis, N., 2015. Insect-specific virus discovery: significance for the arbovirus community. Viruses 7, 4911–4928. https://doi.org/10.3390/v7092851.

Calisher, C.H., Higgs, S., 2018. The discovery of arthropod-specific viruses in hematophagous arthropods: an open door to understanding the mechanisms of arbovirus and arthropod evolution? Annu. Rev. Entomol. 63, 87–113. https://doi.org/10.1146/annurev-ento-020117.

Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., de Matos, A.P., Piedade, J., de Almeida, A.P.G., Parreira, R., 2015. Negeviruses found in multiple species of mosquitoes from southern Portugal: isolation, genetic diversity, and replication in insect cell culture. Virology 483, 318–328. https://doi.org/10.1016/j.virol.2015.04.021.

Colmant, A.M.G., Hobson-Peters, J., Bielefeldt-Ohmann, H., van den Hurk, A.F., Hall- Mendelin, S., Chow, W.K., Johansen, C.A., Fros, J., Simmonds, P., Watterson, D., Cazier, C., Etebari, K., Asgari, S., Schulz, B.L., Beebe, N., Vet, L.J., Piyasena, T.B., Nguyen, H.D., Barnard, R.T., Hall, R.A., 2017. A new clade of insect-specific flaviviruses from Australian Anopheles mosquitoes displays species-specific host restriction. mSphere 2 https://doi.org/10.1128/mSphere.e00262–17.

Cook, S., Moureau, G., Harbach, R.E., Mukwaya, L., Goodger, K., Ssenfuka, F., Gould, E., Holmes, E.C., De Lamballerie, X., Cook, S., 2009. Isolation of a novel species of flavivirus and a new strain of Culex flavivirus (Flaviviridae) from a natural mosquito population in Uganda. Journal of General Virology 2669–2678. https://doi.org/10.1099/vir.0.014183-0.

Cotmore, S.F., Agbandje-McKenna, M., Chiorini, J., Mukha, D.V., Pintel, D.J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D., Davison, A.J.,

2014. The family Parvoviridae. Arch. Virol. 159, 1239–1247. https://doi.org/10.1007/s00705-013-1914-1.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., de Micco, P., de Lamballerie, X., 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J. Gen. Virol. 85, 1971–1980. https://doi.org/10.1099/vir.0.79850-0.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. https://doi.org/10.1186/1471-2148-7-214.

Folmer, O., Black, M., Hoeth, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299. https://doi.org/10.1371/journal.pone.0013102.

Grobbelaar, A.A., Weyer, J., Moolla, N., Van Vuren, P.J., Moises, F., Paweska, J.T., 2016. Resurgence of yellow fever in Angola, 2015–2016. Emerg. Infect. Dis. 22, 1854–1855. https://doi.org/10.3201/eid2210.160818.

Gu, J., Liu, M., Deng, Y., Peng, H., Chen, X., 2011. Development of an efficient recombinant mosquito densovirus-mediated RNA interference system and its preliminary application in mosquito control. PLoS One 6, 1–10. https://doi.org/10.1371/journal.pone.0021329.

Junglen, S., Drosten, C., 2013. Virus discovery and recent insights into virus diversity in arthropods. Curr. Opin. Microbiol. 16, 507–513. https://doi.org/10.1016/j.mib.2013.06.005.

Katoh, K., Standley, D.M., 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.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, 1–5. https://doi.org/10.1093/ve/vev003.

Martynova, E.U., Schal, C., Mukha, D.V., 2016. Effects of recombination on densovirus phylogeny. Arch. Virol. 161, 63–75. https://doi.org/10.1007/s00705-015-2642-5.

Matsuno, K., Weisend, C., Kajihara, M., Matysiak, C., Williamson, B.N., Simuunza, M., Mweene, A.S., Takada, A., Tesh, R.B., Ebihara, H., 2015. Comprehensive molecular detection of tick-borne phleboviruses leads to the retrospective identification of taxonomically unassigned bunyaviruses and the discovery of a novel member of the genus Phlebovirus. J. Virol. 89, 594–604. https://doi.org/10.1128/JVI.02704-14.

Nouri, S., Matsumura, E.E., Kuo, Y.W., Falk, B.W., 2018. Insect-specific viruses: from discovery to potential translational applications. Curr. Opin. Virol. 33, 33–41. https://doi.org/10.1016/j.coviro.2018.07.006.

Pimentel, V., Afonso, R., Nunes, M., Vieira, M.L., Bravo-Barriga, D., Frontera, E., Martinez, M., Pereira, A., Maia, C., Paiva-Cardoso, M. das N., Freitas, F.B., Abecasis, A.B., Parreira, R., 2019. Geographic dispersal and genetic diversity of tick-borne phleboviruses (Phenuiviridae, Phlebovirus) as revealed by the analysis of L segment sequences. Ticks Tick. Borne. Dis. 10, 942–948. https://doi.org/10.1016/j.ttbdis.2019.05.001.

Ratnasingham, S., Hebert, P., 2007. BOLD: the barcode of life data system (www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364. https://doi.org/10.1111/j.1471-8286.2006.01678.x.

Ren, X., Hoiczyk, E., Rasgon, J.L., 2008. Viral paratransgenesis in the malaria vector Anopheles gambiae. PLoS Pathog. 4, 4–11. https://doi.org/10.1371/journal.ppat.1000135.

Sánchez-Seco, M.P., Rosario, D., Quiroz, E., Guzmán, G., Tenorio, A., 2001. A generic nested-RT-PCR followed by sequencing for detection and identification of members of the Alphavirus genus. J. Virol. Methods 95, 153–161. https://doi.org/10.1016/S0166-0934(01)00306-8.

Simon-Loriere, E., Faye, O., Prot, M., Casademont, I., Fall, G., Fernandez-Garcia, M., Diagne, M., Kipela, J.-M., Fall, I., Holmes, E., Sakuntabhai, A., Sall, A., 2017.

Autochthonous Japanese encephalitis with yellow fever coinfection in Africa. N. Engl. J. Med. 376, 1483–1485. https://doi.org/10.1056/NEJMc1701600.

Smith, J.L., Fonseca, D.M., 2004. Rapid assays for identification of members of the Culex (Culex) pipiens complex, their hybrids, and other sibling species (Diptera: Culicidae). Am. J. Trop. Med. Hyg. 70, 339–345.

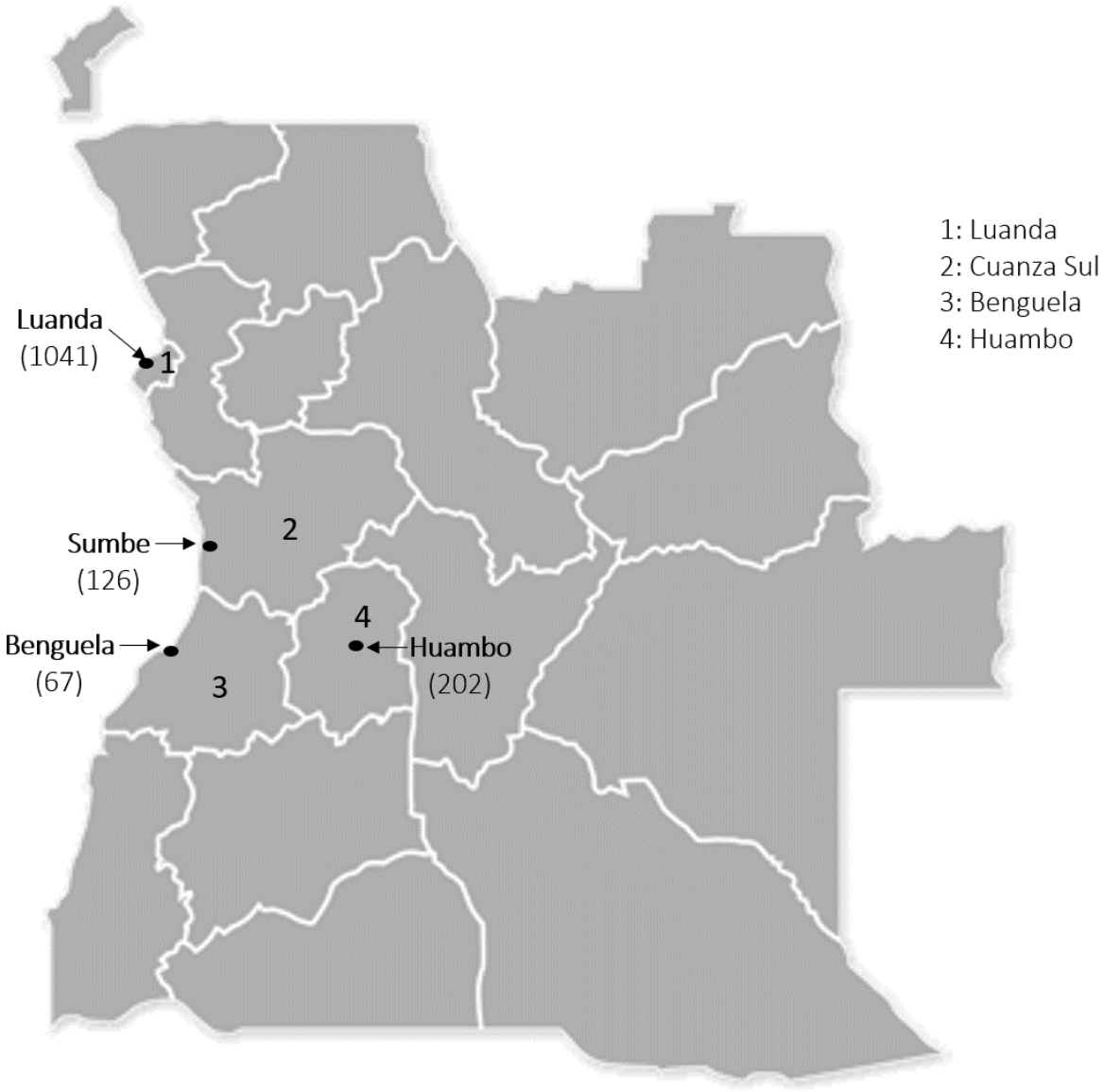
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. https://doi.org/10.1093/molbev/mst197.

Tijssen, P., Pénzes, J.J., Yu, Q., Pham, H.T., Bergoin, M., 2016. Diversity of small, single stranded DNA viruses of invertebrates and their chaotic evolutionary past. J. Invertebr. Pathol. 140, 83–96. https://doi.org/10.1016/j.jip.2016.09.005.

Vázquez, A., Sánchez-Seco, M.-P., Palacios, G., Molero, F., Reyes, N., Ruiz, S., Aranda, C., Marqués, E., Escosa, R., Moreno, J., Figuerola, J., Tenorio, A., 2012. Novel flaviviruses detected in different species of mosquitoes in Spain. Vector-Borne Zoonotic Dis. 12, 223–229. https://doi.org/10.1089/vbz.2011.0687.

Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: Genome detection and phylogenetic characterization of viral sequences

Supplementary Fig. 1: Geographic distribution of the mosquito collection sites. The Angolan provinces where mosquito collections took place are indicated by numbers (legend on the right side of the map), and the approximate location of the main collection sites in each region are identified by an arrow and a dot. For each collection site, the numbers in brackets indicate the total number of mosquitoes collected/analyzed.



Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: Genome detection and phylogenetic characterization of viral sequences

Supplementary Table 1: Primers and thermal profiles used for the detection of densoviruses (*Densovirinae*).

Primer	PCR round	Sequence	Coordinates[a]	Genomic target
VPF1	1st*	GGCAGACAGCACTWCAATGGACC	2567-2589	VP1
VPF2	2nd*	CGTAAAAGAAGGWTAYGGACC	2672-2692	VP1
VPR1	1st*	GTGCTCATTCKTACTTGGTATCTG	3466-3489	VP1
VPR2	2nd*	GTTTCRTCTGGAAYTTGTGGTTGTGC	3425-3450	VP1
DF1	1st**	AACCGTTGGTGACCTCTACCCACAATTAC	737-765	NS1
DF2	2nd**	GAAACTATGGATAATAACGGGTCACAGG	936-963	NS1
DR1	1st**	CGGATGCAATAGAGAATGAAGTTCCTGAG	1590-1618	NS1
DR2	2nd**	CGCTTCTGCACTTCCTGCGCTTGTCGC	1535-1561	NS1

[a] The coordinates indicated correspond to the position of sequences targeted by the different primers in reference FJ805445; K: G or T, R: A or G, W: A or T, Y: C or T.
***** The thermal profile used for amplification of NS1 sequences included (in both PCR cycles) 2'-94 °C (1x), [30''-94 °C, 30''-53 °C, 45''-72 °C (50x)], 5'-72 °C (1x).
****** The thermal profile used for amplification of VP sequences included (in both PCR cycles): 2'-94 °C (1x), [30''-94 °C, 1'-50 °C, 1'-72 °C (50x)], 5'-72 °C (1x).

Supplementary Table 2a: List of viral sequences described in this work.

Viral *taxon*: *Flavivirus*	**Sequence accession number**	**Mosquito pool**
	LC485965[1]	Ang3
	LC485966[1]	Ang4
	LC485967[1]	Ang24
	LC485968[1]	Ang45
Viral *taxon*: *Brevidensovirus*	**Sequence accession number**	**Mosquito pool**
	LC485294[2]	Ang20
	LC485295[2]	Ang22
	LC485296[2]	Ang24
	LC485297[2]	Ang25
	LC485298[2]	Ang27
	LC485299[2]	Ang29
	LC485300[2]	Ang30
	LC485301[2]	Ang31
	LC485302[2]	Ang33
	LC485303[2]	Ang37
	LC485954[3]	Ang20
	LC485955[3]	Ang22
	LC485956[3]	Ang24
	LC485957[3]	Ang25
	LC485958[3]	Ang27
	LC485959[3]	Ang29
	LC485960[3]	Ang30
	LC485961[3]	Ang31
	LC485962[3]	Ang33
	LC485963[3]	Ang37
	LC485964[3]	Ang45

[1] NS5 sequences, [2] VP sequences, [3] NS1 sequences.

Supplementary Table 2b: *Flavivirus* and *Brevidensovirus* detection in mosquitoes from Angola, and characterization of the mosquito pools analysed.

Pool lab code (size)	Collection site	Mosquito sex/Identification*	Flavivirus detection	Brevidensovirus detection
Ang3 (n=25)	Luanda	Male/*Anopheles* sp.	**Positive**	Negative
Ang4 (n=32)	Luanda	Male/*Anopheles* sp.	**Positive**	Negative
Ang5 (n=25)	Luanda	Female/*Anopheles* sp.	**Positive**	Negative
Ang6 (n=24)	Luanda	Male/ND	Negative	Negative
Ang7 (n=14)	Luanda	Female/ND	Negative	Negative
Ang8 (n=25)	Luanda	Male/ND	Negative	Negative
Ang9 (n=25)	Luanda	Male/ND	Negative	Negative
Ang10 (n=24)	Luanda	Male/ND	Negative	Negative
Ang11 (n=25)	Luanda	Female/ND	Negative	Negative
Ang12 (n=17)	Luanda	Female/ND	Negative	Negative
Ang13 (n=20)	Luanda	Male/ND	Negative	Negative
Ang14 (n=30)	Luanda	Female/ND	Negative	Negative
Ang15 (n=28)	Luanda	Female/ND	Negative	Negative
Ang16 (n=10)	Luanda	Male/ND	Negative	Negative
Ang17 (n=21)	Luanda	Female/ND	Negative	Negative
Ang18 (n=12)	Luanda	Male/ND	Negative	Negative
Ang19 (n=9)	Luanda	Female/ND	Negative	Negative
Ang20 (n=25)	Luanda	Female/*Culex pipiens* s.l.	Negative	**Positive**
Ang21 (n=25)	Luanda	Female/ND	Negative	Negative
Ang22 (n=30)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang23 (n=30)	Luanda	Female/ND	Negative	Negative
Ang24 (n=30)	Luanda	Male/*Culex pipiens* s.l.	**Positive**	**Positive**
Ang25 (n=25)	Luanda	Female/*Culex pipiens* s.l.	Negative	**Positive**
Ang26 (n=16)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang27 (n=30)	Luanda	Female/*Culex pipiens* s.l.	Negative	**Positive**
Ang28 (n=30)	Luanda	Female/*Culex pipiens* s.l.	Negative	**Positive**
Ang29 (n=31)	Luanda	Female/*Culex pipiens* s.l.	Negative	**Positive**
Ang30 (n=30)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang31 (n=21)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang32 (n=52)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang33 (n=50)	Luanda	Female/*Culex pipiens* s.l.	Negative	**Positive**
Ang34 (n=34)	Luanda	Female/ND	Negative	Negative
Ang35 (n=50)	Luanda	Female/ND	Negative	Negative
Ang36 (n=55)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang37 (n=30)	Luanda	Male/*Culex pipiens* s.l.	Negative	**Positive**
Ang38 (n=25)	Benguela	Female/ND	Negative	Negative
Ang39 (n=46)	Huambo	Female/ND	Negative	Negative
Ang40 (n=36)	Huambo	Female/ND	Negative	Negative
Ang41 (n=11)	Sumbe	Female/ND	Negative	Negative
Ang42 (n=49)	Huambo	Male/ND	Negative	Negative
Ang43 (n=14)	Luanda	Male/ND	Negative	Negative
Ang44 (n=25)	Benguela	Female/ND	Negative	Negative
Ang45 (n=25)	Sumbe	Female/*Culex pipiens* s.l.	**Positive**	**Positive**
Ang46 (n=25)	Huambo	Female/ND	Negative	Negative
Ang47 (n=9)	Luanda	Male/ND	Negative	Negative
Ang48 (n=7)	Luanda	Female/ND	Negative	Negative
Ang49 (n=16)	Luanda	Male/ND	Negative	Negative
Ang50 (n=18)	Luanda	Female/ND	Negative	Negative
Ang51 (n=13)	Luanda	Male/ND	Negative	Negative
Ang52 (n=17)	Sumbe	Female/ND	Negative	Negative
Ang53 (n=27)	Sumbe	Female/ND	Negative	Negative
Ang54 (n=19)	Sumbe	Male/ND	Negative	Negative
Ang55 (n=23)	Huambo	Female/ND	Negative	Negative
Ang56 (n=10)	Huambo	Female/ND	Negative	Negative
Ang57 (n=13)	Huambo	Male/ND	Negative	Negative
Ang58 (n=9)	Sumbe	Male/ND	Negative	Negative
Ang59 (n=18)	Sumbe	Female/ND	Negative	Negative
Ang60 (n=17)	Benguela	Female/ND	Negative	Negative
Ang61 (n=2)	Luanda	Female/ND	Negative	Negative
Ang62 (n=2)	Luanda	Female/ND	Negative	Negative
60 pools (n=1,436 mosquitoes)			n=5 positive	n=15 positive

*Identification was based on COI barcoding using sequences generated by Sanger (population) sequencing of PCR products.

Supplementary Table 3: Estimates of evolutionary divergence over sequence pairs between groups

		1	*2*	*3*	*4*	*5*
1	*Brevidensovirus*	--	2.65/2.56	3.29/2.17	2.21/2.51	1.68/2.76
2	*Ambidensovirus*	2.65/2.56	--	2.51/2.13	2.77/2.85	2.64/3.39
3	*Iteradensovirus*	3.29/2.17	2.51/2.13	--	5.73/3.01	4.32/2.64
4	*Hepandensovirus*	2.21/2.51	2.77/2.85	5.73/3.01	--	8.44/3.05
5	*Penstyldensovirus*	1.68/2.76	2.64/3.39	4.32/2.64	8.44/3.05	--

The values indicate amino acid substitutions per site from averaging over all sequence pairs between the defined groups (genera). All ambiguous positions were removed from each sequence pair.

Chapter 4. Genetic lineage characterization and spatiotemporal dynamics of classical insect-specific flaviviruses: outcomes and limitations

Published as:

Morais, P., Trovão, N. S., Abecasis, A. B., & Parreira, R. (2021). Genetic lineage characterization and spatiotemporal dynamics of classical insect-specific flaviviruses: outcomes and limitations. Virus Research, 303 (October), 198507. https://doi.org/10.1016/j.virusres.2021.198507

Abstract

The genus Flavivirus incorporates bona fide arboviruses, as well as others viruses with restricted replication in insect cells. Among the latter, a large monophyletic cluster of viruses, known as cISF (classical insect-specific flaviviruses), has been sampled in many species of mosquitoes collected over a large geographic range.

In this study, we investigated nucleotide and protein sequences with a suite of molecular characterization approaches including genetic distance, Shannon entropy, selective pressure analysis, polymorphism identification, principal coordinate analysis, likelihood mapping, phylodynamic reconstruction, and spatiotemporal dispersal, to further characterize this diverse group of insect-viruses. The different lineages and sub-lineages of viral sequences presented low sequence diversity and entropy (though some displayed lineage-specific polymorphisms), did not show evidence of frequent recombination and evolved under strong purifying selection. Moreover, the reconstruction of the evolutionary history and spatiotemporal dispersal was highly impacted by overall low signals of sequence divergence throughout time but suggested that cISF distribution in space and time is dynamic and may be dependent on human activities, including commercial trading and traveling.

Keywords: Flavivirus; Insect-specific viruses; Genetic diversity; Phylogenetic analysis; Phylogeography; Phylodynamics; BEAST

1. Introduction

The genus *Flavivirus* (*Flaviviridae*) encompasses a genetically diverse array of enveloped RNA viruses with single-stranded genomes of positive polarity, many of which are well-known tick- or mosquito-borne pathogenic arboviruses (Holbrook, 2017). Many of them have (re-)emerged in recent years to become worldwide threats causing millions of infections on a global scale (Bhatt et al., 2013; Chong et al., 2019), while others (e.g. the Japanese encephalitis and yellow fever viruses) still have the potential to expand their geographical distribution and increase their current burden on human health (Wasserman et al., 2016; Weaver et al., 2018).

Although many flaviviruses are pathogenic to humans and some (eg. the Kyasanur forest and Omsk hemorrhagic fever viruses) are studied under strict biological containment (Wilson and Chosewood, 2009), the replication of others seems to be restricted to invertebrate hosts (Blitvich and Firth, 2015). These, commonly designated insect-specific flaviviruses (ISF), are divided into two groups. One includes the so-called dual host-affiliated ISF (dhISF) or lineage II ISF (Harrisom et al., 2020). They do not form a monophyletic lineage in the flavivirus phylogenetic tree, but cluster among mosquito-borne arboviruses and their restriction to mosquitoes appears to have resulted from a secondary loss of their ability to replicate in vertebrates (Blitvich and Firth, 2015). On the other hand, most ISF described to date cluster in a separate large monophyletic cluster (lineage I), and are commonly known as classical ISF (cISF), a group that includes the first of these viruses (the cell fusing-agent virus) described more than 45 years ago (Stollar and Thomas, 1975).

From the moment the study of ISF was revived in the early 2000s with the characterization of the Kamiti river virus, isolated from *Aedes macintoshi* collected in Kenya (Crabtree et al., 2003; Sang et al., 2003), many others have been described, indicating that these viruses not only have global distribution but can also be associated with both Culicine and Anopheline mosquitoes (Calzolari et al., 2016; Colmant et al., 2017). They have been suggested to correspond to an ancestral lineage of flaviviruses (Cook et al., 2012) but their evolutionary history is unclear and complex, and can even be found as endogenous viral elements in the genome of mosquitoes (Abílio et al., 2020; Crochu et al., 2004; Nouri et al., 2018). Furthermore, the study of ISF has gained

increasing momentum in more recent years as translational opportunities involving their potential use as biological agents to interfere with vector competence (via superinfection exclusion) have been explored (Goenaga et al., 2020; Hobson-Peters et al., 2013; Kuwata et al., 2015; Romo et al., 2018).

Nonetheless, the great majority of the studies involving ISF include their identification, genetic characterization and phylogenetic placement within the genus *Flavivirus*. Very rarely have questions such as the evolutionary and ecological hypotheses ruling their origin and spatiotemporal dispersal been addressed, and in the few studies where these analyses have been carried out (Cella et al., 2019), no coherent statistical framework was used. For this reason, we attempted to do so by focusing on some of the most representative genetic sublineages found within the cISF radiation using either non-structural protein 5 (or ns5, encoding the viral RNA dependent RNA polymerase, and the most frequently used markers used for genetic analyses of cISF) or complete genome nucleotide sequences available at public databases. Different bioinformatics tools were employed to analyze each of these groups of sequences, and the obtained reconstructions demonstrate the value of a Bayesian-based phylodynamic model to infer the evolutionary history and spatial spread of different subgroups of cISF.

2. Materials and Methods

2.1. Dataset preparation and sequence alignments

The compilation of the different nucleotide (nt) sequence datasets (ds) used in this work was based on the selection of non-structural protein 5 (NS5) and whole-genome sequences available in GenBank on 31-April-2020. These were either directly identified via their accession numbers or indirectly singled-out as a result of similarity searches using BLASTn. For those viruses for which a complete genomic sequence was available, protein-datasets (including non-structural and structural proteins) were also assembled.

Our analyses focused on eight datasets (ds) that include nucelotide sequences clustering as major branches on the phylogenetic tree as depicted in Supplementary Figure 1. These included partial *Culex theileri* flavivirus (n=80; ds1), *Culex pipiens* flavivirus (n=41; ds2), *Culex* flavivirus (n=172; ds3), *Aedes* flavivirus (n=59; ds4), cell-fusing agent virus

(CFAV) (n=41; ds5), and cISF NS5 coding sequences, the latter including a maximum of two sequences from the same genetic lineage per country per year of sampling (n=95; ds6). Two complete Open Reading Frame (ORF)-coding datasets were also analysed, one including *Culex* flavivirus sequences (n=45; ds7), and the other cISF sequences, the latter comprising a maximum of two sequences from the same genetic lineage per country per year of sampling (n=83; ds8).

Multiple alignments of nt sequences were performed using the iterative G-INS-I method as implemented in MAFFT vs. 7 (Katoh and Standley, 2013), followed by their edition using GBlocks (Castresana, 2000), allowing for less strict flanking positions. The multiple sequence alignments were systematically verified to ensure the correct alignment of homologous codons using BioEdit 7.0.5 (Hall, 1999).

2.2. Assessment of the temporal and phylogenetic signals of different sequence datasets

Inspection of the degree of temporal signal (i.e. signal for divergence accumulation over the sampling time interval) was carried out based on an exploratory linear regression approach using the topology of a maximum likelihood (ML) tree, estimated under an unconstrained clock and GTR+Γ+I substitution model using the MEGA X software (Kumar et al., 2018). Root-to-tip divergences were plotted as a function of sampling time using the TempEst software (Rambaut et al., 2016).

On the other hand, the evolutionary information contained in each aligned dataset (phylogenetic signal) was assessed by likelihood mapping (Strimmer and von Haeseler, 1997) using TREE-PUZZLE v5.3 (Schmidt et al., 2002).

2.3. Phylogenetic analyses using maximum likelihood and Bayesian approaches

Phylogenetic reconstructions were performed based on the ML optimization criterion using the GTR+Γ+I as the best data-fitting substitution model, as suggested by IQ-TREE (Trifinopoulos et al., 2016), and the stability of the obtained tree topologies was assessed

by bootstrapping with 1000 re-samplings of the original sequence data in MEGA X software.

Time-calibrated phylogenetic and phylogeographic histories were constructed using a Bayesian statistical framework implemented in the BEAST v1.10 software package (Suchard et al., 2018), using GTR+Γ+I model. All phylogenetic reconstructions were carried out assuming a relaxed uncorrelated lognormal molecular clock model (Ho et al., 2005) as indicated by the ML Clock Test implemented in MEGA X, allowing for the accommodation of among-lineage rate variation.

To investigate the sensitivity of the estimate for the time to the Most Recent Common Ancestor (tMRCA) concerning the coalescent priors used, the performance of parametric demographic priors (constant, exponential, logistic, and expansion population growth priors) (Drummond et al., 2003; Griffiths and Tavare, 1994) were tested against that of nonparametric ones (Bayesian Gaussian Markov Random Field (GMRF) Skyride (Minin et al., 2008), Skygrid (Gill et al., 2013) and Skyline (Drummond et al., 2005)). This preliminary comparative analysis was carried out using two datasets of partial *ns5* sequences including *Culex theileri* and *Culex pipiens* cISF, while the performance of nonparametric demographic priors was also estimated using a *ns5* dataset of CFAV sequences. Bayes factor (BF) support for predictors was calculated using the marginal likelihood estimates (infered using Path Sampling (PS) and Stepping-Stone (SS) approaches) for each candidate model and then comparing the ratio of the marginal likelihood estimates for the set of candidate models.

A minimum number of two (up to a maximum of eight) independent Markov chain Monte-Carlo (MCMC) runs were performed using BEAST v1.10 until 1-3×108 states were sampled, and at least 10% of which were discarded as burn-in. The length of the MCMC analyses was defined as a function of chain convergence using the Tracer software v1.7.1 (http://beast.bio.ed.ac.uk/tracer), which was also used to check for adequate effective sample size (ESS) higher than 200 after the removal of burn-in. The trees were logged on every 10,000th MCMC step, and the trees distribution was summarized using TreeAnnotator software v1.8.3 as a maximum clade credibility (MCC) tree, using median heights as the node heights in the tree. The FigTree v1.4.2 software was used to visualize the phylogenetic trees (http://tree.bio.ed.ac.uk/software/figtree/).

2.4. Continuous phylogeography

The geographic spread of cISF in continuous space was studied using a phylogenetic Brownian diffusion approach that models the change in geographic coordinates (latitude and longitude) along each branch in the phylogenetic reconstruction (Lemey et al., 2010). As an alternative to homogeneous Brownian motion, relaxed random walk (RRW) extensions that model branch-specific variation in dispersal rates similar to uncorrelated relaxed clock approaches was also used (Drummond et al., 2006a). The assessment of BF support for the diffusion priors was calculated as described above for the coalescent demographic priors.

The spatiotemporal reconstruction of ISF spread was visualized on the Spatial Phylogenetic Reconstruction of Evolutionary Dynamics software (SpreaD3; Bielejec et al., 2016), using a custom-made geoJSON world map (https://geojson-maps.ash.ms/).

2.5. Genetic diversity and selective pressure analyses

The putative mosaic structure of cISF sequences was investigated using the Recombination Detection Program RDP4 (Martin et al., 2015), and the estimation of genetic distance values (corrected with the Kimura-2P formula) was carried out using MEGA X. Single amino-acid polymorphisms (SAPs) for protein variation were detected with the indicated amino acid coordinates corresponding to those in the CFAV reference sequence NC_001564.

The analyses of selective pressure on individual sites of codon alignments were carried out using the Single Likelihood Ancestor Counting (SLAC) and Fixed Effects Likelihood (FEL) methods as implemented in Datamonkey (Kosakovsky Pond and Frost, 2005), or using the SNAP tool (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP. html) that explores a simple method for calculation of synonymous and non-synonymous substitutions (Nei and Gojoborit, 1986). The degree of variability of each amino acid position in multiple alignments of the putative ISF NS5 amino acid sequences evaluated based on the Shannon entropy function was calculated using Entropy (Shannon entropy-one and entropy-two options; available at http://www.hiv.lanl.gov/content/sequence/ENTROPY/ entropy. html). Finally, a

principal coordinate analysis was carried out using PCOORD (http://www.hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html).

3. Results

3.1. Genetic diversity and selective pressure analyses

The non-structural protein 5 viral sequences of *Culex*-associated *Culex theileri* flaviviruses (CTFV) and *Culex pipiens* flaviviruses (CPFV), as well as cell-fusing agent viruses (CFAV), comprised datasets ds1, ds2, and ds5, respectively. In addition to these, three other datasets of more divergent assemblages of *ns5* sequences included *Culex* cISF (ds3) and *Aedes* cISF (ds4), as well as the whole cISF radiation (ds6). Finally, we also set up two datasets of complete viral genomic ORF-coding sequences, which integrated *Culex*-specific cISF (ds7) or representative sequences of all cISF (ds8). Access to complete genomic sequences also allowed the construction of multiple datasets for each of the non-structural (NS1, NS2a, NS2b, NS3, NS4a, 2k, NS4b, NS5) as well as structural (capsid, envelope, and membrane glycoprotein) protein sequences.

Preliminary ML phylogenetic tree of all *Culex* cISF partial NS5 coding sequences (ds3) identified multiple sublineages (indicated as "L" in Supplementary Figure 2A), most of them associated with a specific species of *Culex* mosquitoes (for example, L1-*Culex theileri*; L2-*Culex tritaeniorhynchus*; L3-*Culex pipiens*). The existence of genetic sublineages was also supported by PCOORD analysis (Supplementary Figure 2B). Curiously, one of the sequences (accession number LC462017), detected in *Culex antennatus* from Mozambique was placed closer to *Culex tritaeniorhynchus* cISF (L2) than to those it forms a sublineage with (L4, including Nienokoue viral sequences). Shannon entropy assessment showed low values for different sublineages of *Culex* cISF (data not shown).

Overall mean genetic distance for *Culex* cISF was calculated for both the complete genome and for each specific genomic region. The mean distance between all genomic sequences was 0.302, with the lowest and highest values (0.252 and 0.399, respectively) associated with the NS5 and C protein-coding sequences, respectively (Supplementary Figure 3A). Not surprisingly, the inclusion of a more divergent group of ORF-coding

sequences into a single dataset raised the average genetic distances to 0.391, with the lowest and highest values (0.327 and 0.498, respectively) associated with the NS5 and NS2b protein-coding sequences, respectively (Supplementary Figure 3B).

As it has been described for *Culex* cISF, a preliminary reconstruction of the phylogenetic relationships of all *Aedes* cISF partial NS5 coding sequences identified three different sublineages, one of them being CFAV (Supplementary Figure 4A). Mean *ns5* genetic distance was, as expected, higher for *Aedes* cISF (encompassing viral sequences found in multiple species of *Aedes* mosquitoes) than CFAV (0.14 and 0.03, respectively), only found to be associated with *Aedes aegypti* and corresponding to a single genetic lineage. Even though low diversity between CFAV was apparent, PCOORD did suggest different subclusters of sequences even in the CFAV sublineage (Supplementary Figure 4B).

Multiple single amino-acid polymorphisms for protein variation were detected in the *ns5* protein between different cISF sublineages. Most SAPs such as 2949G, 2986S, 3014M, 3068A, 3073S, 3117S, 3172E, and 3195M were found to be characteristic of *Aedes* cISF sequences, while polymorphism 3044R was found in CFAV sequences but not in other *Aedes* cISF. Multiple SAPs were also only found on *Culex* cISF, including 3000Q, 3017C, 3083F, and 3193G, with 3094G and 3204G only being found in *Cx. theileri* cISF.

Estimation of omega (ω) values (corresponding to the ratio of non-synonymous to synonymous substitutions) was performed for cISF using three different methods (SLAC, FEL, and SNAP) based on the analysis of both the complete genome and for each genomic region (Supplementary Table 1). Overall results indicate that the whole genome is under strong purifying selection, with low ω values, as previously suggested for CTFV *ns5* (Cella et al., 2019). Most structural proteins and non-structural proteins possess little to nil evidence of positively selected sites. On the contrary, smaller genomic regions seemed to be under lower levels of purifying selection, with the *ns2* region presenting the highest ω, and a higher percentage of positively selected sites, including codons 921, 925, 929, and 936. *Culex* and *Aedes*-specific cISF selective pressure analyses disclosed similar results (Supplementary Table 2).

Using as a reference a minimum of 84% of identity between *ns5* nucleotide sequences to define a viral species (Kuno et al., 1998; Peterson, 2014), calculation of genetic diversity values supported the delimitation of independent viral lineages mostly associated with a

single mosquito species, some of which corresponded to major branches on a phylogenetic tree (Supplementary Figure 1), defining, for example, *Cx. theileri* (CT) and *Cx. pipiens* (CP) flaviviruses (CTFV and CPFV, respectively), or the *Aedes aegypti*-associated cell-fusing agent virus (or CFAV). Phylogenetic reconstructions did not seem to be impacted by the occurrence of pervasive intra- or inter-lineage recombination events. Nonetheless, we detected a possible recombination event among Calbertado virus sequences. RDP4 indicated that viral sequences with accession numbers KX669683 and KX669685 could stem from recombination between the *ns1*, *ns2a,* and *ns2b* genomic regions of KX669686 (Calbertado virus from Canada) and the remaining coding sequence of KX669682 (Calbertado virus from EUA) (data not shown). Consequently, these sequences were excluded from phylogenetic reconstructions that involved the analysis of complete ORF-coding sequences.

3.2. Assessment of phylogenetic signal and analysis of sequence divergence throughout time

Phylogenetic signal was evaluated for each nt dataset using likelihood mapping. The obtained results (Table 1) showed that the percentage of totally resolved sequence quartets (of the total number of their possible number in 10,000 replicates) ranged from 89.6% to 99.6%, with the highest values associated with datasets including *Culex*-specific cISF (ds7) or representative sequences of all cISF (ds8), with 99.6% unambiguous quartet resolution in both cases, the latter corresponding to assemblages of complete ORF-coding sequences. Given the overall high phylogenetic signal for all the eight datasets, it is not surprising that most of the major branches in phylogenetic reconstructions of the genus *Flavivirus* were found to be topologically sound. Additional likelihood mapping analyses were also executed for each genomic region for complete genome sequence datasets (Supplementary Table 3). For both *Culex*-specific cISF and all cISF, *ns3*- and *ns5*-specific datasets displayed higher phylogenetic signals, with both phylogenetic trees showing equivalent topologies to those reconstructed with complete ORF-coding sequences (data not shown).

To assess the extent to which all datasets contained detectable signals for sequence divergence throughout time, a standard linear regression exploration of root-to-tip genetic

distances as a function of sampling time was performed. Except for datasets ds1 and ds5 (CTFV and CFAV, respectively), the remainder did not reveal clear evidence for a substantial temporal signal (Table 1). This suggests that regardless of their overall high phylogenetic signal, either the temporal signal estimate for the different datasets may be strongly impacted by nucleotide substitution rate variation, especially among the deeper tree branches, or that that the interval of dates of sampling is not broad enough. As far as the first possibility is concerned, whereas the rates of evolutionary change have already been estimated for various pathogenic flaviviruses (Araújo et al., 2009; Barzilai and Schrago, 2019a; Sall et al., 2010; Sang et al., 2019a), the same does not apply to cISF.

Table 1: Phylogenetic signal (as assessed by likelihood mapping) and root-to-tip (sequence divergence as a function of time) of cISF sequences using datasets of *ns5* or complete ORF-coding sequences.

	Datasets							
	ns5						**complete ORF**	
Likelihood Mapping	**ds1**	**ds2**	**ds3**	**ds4**	**ds5**	**ds6**	**ds7**	**ds8**
Totally resolved quartets	89.8%	89.6%	96.5%	88.6%	90.3%	98.4%	99.6%	99.6%
Partially resolved quartets	2.9%	3.1%	1.6%	4.8%	3.5%	1.2%	0.3%	0.3%
Unresolved quartets	7.3%	7.3%	2.0%	6.5%	6.1%	0.4%	0.1%	0.1%
Root-to-tip analysis (r^2)	0.347	0.022	$4.6x10^{-4}$	0.046	0.534	0.085	0.019	0.0096

partial NS5-coding sequences: ds1 - *Culex theileri* flavivirus (n=80); ds2 - *Culex pipiens* flavivirus (n=41); ds3 - *Culex* flavivirus (n=172); ds4 - *Aedes* flavivirus (n=59); ds5 - cell fusing agent virus (n=41); ds6 - partial cISF (n=95). Complete ORF-coding sequences: ds7 - *Culex* flavivirus (n=45); ds8 - cISF (n=83).

Therefore, nucleotide substitution rates were estimated using the sequences of the three different datasets under analysis defining viral lineages associated with a specific mosquito species (CTFV – ds1; CPFV – ds2; and CFAV – ds5), and assuming a relaxed molecular clock model (Drummond et al., 2006b). This assumption was supported both by the Maximum Likelihood test of the molecular clock hypothesis, performed for each dataset, and which systematically rejected the null hypothesis of equal nucleotide substitution rates along the branches of the trees, and the exploratory linear regression analysis results, indicating that for most datasets there was no apparent time-dependent accumulation of divergence (from the tips of branches to the root of the tree). Curiously, while the molecular clock test systematically rejected the strict-clock hypothesis favoring

a relaxed clock, a substantial variation of the substitution rate along the branches of the phylogenetic trees, indicated by the high coefficient of variation of its estimates (Table 2), was not always observed, although it was consistently high for CTFV. The obtained results (Supplementary Table 4) differed according to the coalescent priors used and ranged from 7.12×10^{5} to 1.13×10^{3} substitutions per site/year, although the great majority of them fell in the range of 3.34×10^{4} to 9.88×10^{4} substitutions per site/year. These values are similar to those previously calculated for other flaviviruses including dengue viruses 3 (Araújo et al., 2009) and 4 (Sang et al., 2019b) (8.9×10^{4} and 9.8×10^{4} substitutions per site/year, respectively), the yellow fever virus [approximately 2×10^{4} substitutions per site/year; (Sall et al., 2010)], or the Zika virus [approximately 8×10^{4} substitutions per site/year; (Barzilai and Schrago, 2019b)].

Table 2: Evaluation of rates for coalescent combined with different geographic diffusion priors: analysis of CTFV (ds1), CPFV (ds2) and CFAV (ds5) *ns5* sequences.

	root_age [95% HPD]	**mean rate [95% HPD]**	**coefficient of variation [95% HPD]**
ds1			
Skyline	1966 [1926–1998]	1.13E-03 [6.45E-04–1.67E-03]	1.03 [0.62–1.51]
Skyline+CauchyRRW	1964 [1923–1998]	1.15E-03 [6.81E-04–1.64E-03]	1.07 [0.69–1.51]
ds2			
Skygrid	1717 [1558–1965]	4.62E-04 [9.73E-05–8.49E-04]	0.37 [2.02E-05–0.85]
Skygrid+CauchyRRW	1908 [1799 - 1987]	8.58E-04 [2.73E-04 – 1.51E-03]	0.902 [0.016-1.71]
ds5			
GMRF Skyride	1865 [1767–1984]	3.24E-04 [5.80E-05–6.27E-04]	1.54E-03 [0–1.94E-15]
GMRF Skyride+ CauchyRRW	-65696 [1796–1987]	3.8E-04 [6.82E-05–6.94E-04]	0.0538 [0–0.46]

GMRF: gaussian Markov random field; HPD: highest Probability Density; nd: not determined; PS: path sampling; RRW: Relaxed Random Walk; SS: stepping-stone sampling. The values 1 and 2 associated with PS and SS indicate those obtained in two independent MCMC runs.

3.3. Continuous phylogeography

To attempt to infer the population dynamics of ISF, the performance of parametric demographic priors was tested against that of non-parametric ones using two selected datasets with reasonable (CTFV – ds1; r2=0.347), as well as poor (CPFV – ds2; r2=0.022) temporal signals. Bayes factor (BF), as well as adequate ESS values, clearly revealed better performance for non-parametric demographic priors in both datasets. Interestingly, the best nonparametric prior (Table 2) was not the same for every dataset, with the best candidate models consisting of Bayesian Skyline for CTFV and Bayesian Skygrid for CPFV. On the other hand, since the performance of non-parametric priors was consistently better than that parametric ones (as judged by marginal likelihood and ESS values) when cISF datasets including CTFV and CPFV sequences were analysed, only the performance of the former were evaluated for CFAV (which was characterized by an adequate temporal signal with an r2=0.534 in the root-to-tip analysis; Table 1), and the obtained results pointed towards GMRF Bayesian Skyride as the coalescent prior of choice. This occurred not because of marginal likelihood estimates / BF values, but as a consequence of the fact that the convergence results and ESS values were not reasonable for other parametric priors (for a maximum of 10 independent runs), unlike those obtained with the GMRF Bayesian Skyride model (data not shown).

For the evaluation of what would be the best geographic diffusion model to be used for spatiotemporal dispersal analysis, a comparative assessment of the performance of a strict Brownian vs. several RRW diffusion models was also performed. For this, marginal likelihood estimation was based on the analyses of CTFV and CFAV (Supplementary Table 4). For CTFV BF values suggest that a Cauchy-RRW approach was the best fit to explain its dispersal dynamics. On the other hand, analysis of ds5 (CFAV) revealed that the Cauchy-RRW prior provided the best convergence and ESS values, even though BF values slightly favored Gamma-RRW. In conclusion, for a spatiotemporal dispersal analysis of CTFV and CFAV in continuous space, a Cauchy-RRW diffusion process was combined with either a Bayesian Skyline (CTFV – ds1) or a GMRF Skyride (CFAV – ds5) coalescent prior.

Analyses of the spatiotemporal patterns of viral sequence spread were carried out under the selected priors, and the obtained results graphically described both as MCC

phylogenetic trees (Supplementary Figures 5 and 6) and as visual reconstructions using the SpreaD3 software (Figure 1A and B). These analyses focused on the dispersal of the cISF *ns5* sequences for which an appropriate temporal signal had been calculated, i.e. CTFV (Supplementary Figure 5) and CFAV (Supplementary Figure 6).

The spatiotemporal analysis of the *Culex theileri* flavivirus (CTFV) sequences suggested an unexpectedly recent expansion of CTFV and proposed the existence of a MRCA for those comprising CTFV dating around the late 1960s (95% Highest Probability Density (HPD): 1921- 1998) located in the east Mediterranean region (the suggested origin was Lebanon). From there, two viral lineages seemed to have spread in both eastern and western directions. Viral expansion of the eastern-bound lineage (starting from the MRCA) may have reached South-East Asia (including Myanmar and neighboring countries) in the early 2000s (95% HPD: 1996-2007). The other route (for which an Italian geographical origin was suggested, though not strongly supported by a high posterior probability (PP) (PP=0.08)) seemed to have split into two different routes, most probably reaching Spain in the early 1990s (95% HPD: 1983-1999) and Turkey in the late 2000s (95% HPD: 2005-2011) (Supplementary Figure 5 A1). Interestingly, when the spatiotemporal analysis was restricted to only CTFV-*ns5* sequences with a European origin, the proposed MRCA origin was not Italy, but rather Spain, yet again not strongly supported (PP=0.08) (Supplementary Figure 5 A2). More recent years marked the expansion of the Spanish clade in the country and the movement of CTFV towards Portugal. On the other hand, the available sequence sampling did not indicate a recent expansion of the Turkish lineage beyond Turkey.

A similar investigation of the spatiotemporal patterns of viral sequence dispersal was used to analyze the cell-fusing agent virus (CFAV – ds5) sequence dataset. Similar to CTFV, a recent origin for the MRCA of CFAV was estimated in the late 1930s (95% HPD: 1796-1987). Two viral lineages seemed to have then spread into distinct directions, one outbound to the American continent and another into Southeast Asia, both reaching their destinations in the late 1940s (95% HPD: 1827- 1993). The Asian clade seemed to have

continued

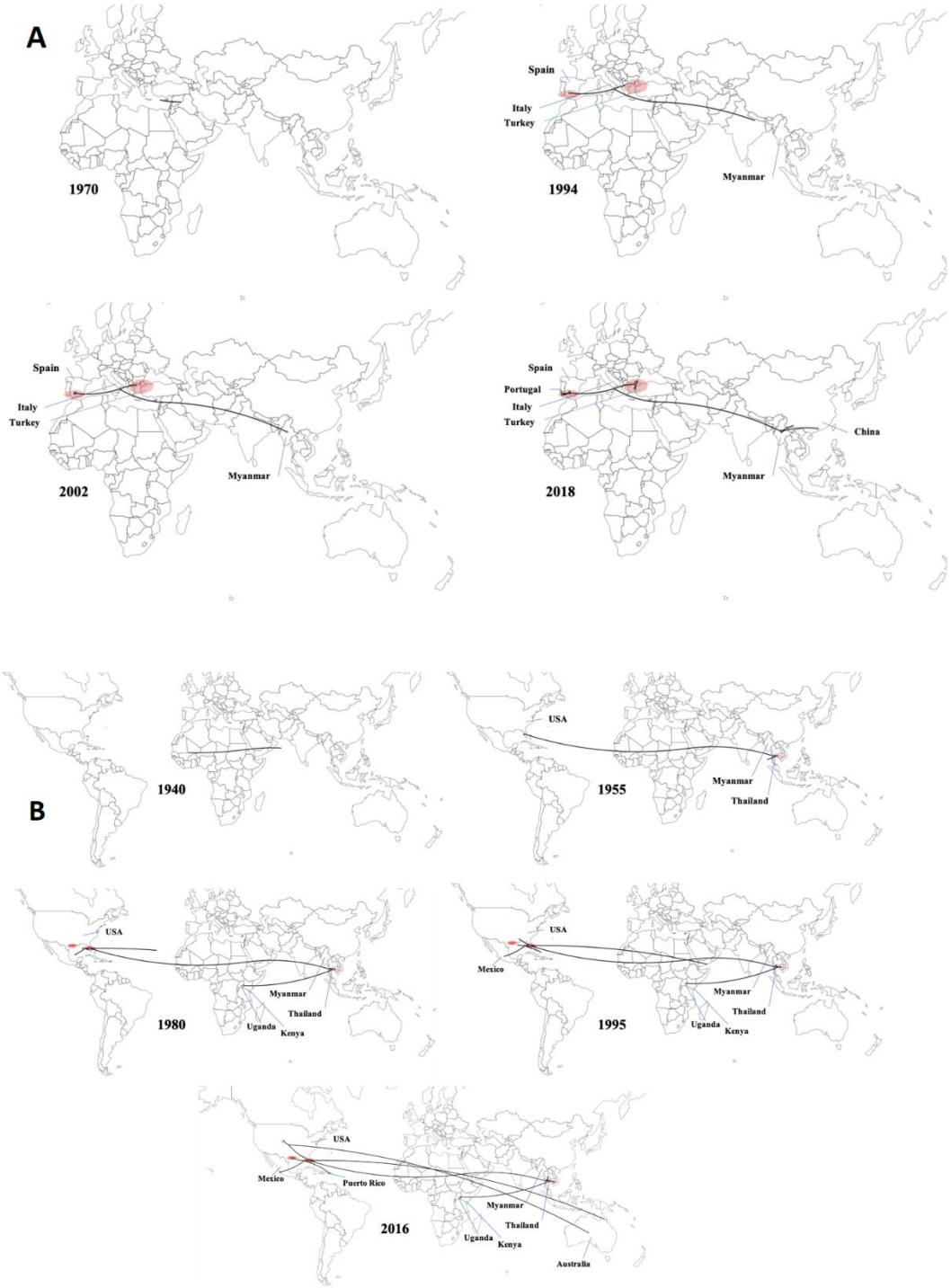


Fig. 1: (A) Spatiotemporal reconstruction of Culex theileri cISF spread visualized on SpreaD3 software, based on the MCC tree represented in Supplementary Figure 5. (B) Spatiotemporal reconstruction of CFAV spread visualized on SpreaD3 software using the MCC tree represented in Supplementary Figure 6.

expanding into Thailand in the late 1980s (95% HPD: 1944-2002). Our analysis also suggested the possibility of one additional expansion event from Southeast Asia (possibly Thailand) to East Africa (Uganda), from where the virus may have later reached neighboring Kenya. The American clade mainly expanded into North America (USA and Mexico), with two subsequent, likely independent, expansion events into Australia more recently (Supplementary Figure 6). Finally, a reconstruction of the continual dispersal of *Culex pipens* cISF (CPFV) was also carried out using the Bayesian Skygrid and Cauchy-RRW prior combination (Supplementary Figure 7). While the suggested age of the MRCA (1916, 95% HPD: 1786-1985) may have been affected by the lack of a strong temporal signal, most of the nodes of the obtained MCC tree were strongly supported, and an Asian origin (China; PP=0.19) was suggested. From here, two major Asian lineages expanded towards Japan (1989, 95% HPD: 1970-2000), Africa (Liberia; 1993, 95% HPD: 1970-2004), and the USA (2000, 95% HPD: 1990- 2005) in more recent times.

4. Discussion

Insect-specific flaviviruses are a group of viruses that infect a wide range of mosquito hosts and display a global geographical distribution among all sublineages (Blitvich and Firth, 2015). In the last decades, new information about their genetics and biological applications (including their potential use as vaccine or heterologous protein expression vectors) has emerged (Holbrook, 2017), but many aspects remain unknown or under present investigation, including their insect-specific viral replication, evolution, dispersal over time, and relationship with other flaviviruses. Our present study hopes to gain information on most, if not all, of these very important aspects.

When phylogenetic relationships among flaviviruses are reconstructed based on the analysis of either *ns5* or whole-genome (ORF-coding) sequences, viral sequences split into two major genetic lineages. One of these incorporates *bone fide* tick- and mosquito-borne arboviruses, dhISF, as well as flaviviruses with no known vector (NKV), whereas the other groups all known cISF. These cISF do not always segregate geographically, but rather seem to associate with either a single species or a small group of species, within a given genus of mosquitoes. Given their medical importance, the genera *Culex*, *Aedes*,

and *Anopheles* are those associated with the majority of currently described cISF, but this most probably results from sampling bias by the more frequent analysis of medically important mosquitoes. In fact, different lineages of these viruses have been detected in *Culiseta* and especially in African *Mansonia* (Abílio et al., 2020).

Arboviruses and cISFs have been shown to share a common evolutionary history, which led to the suggestions that cISFs may eventually bypass host range restrictions imposed by cells in vertebrate hosts (Junglen et al., 2017) and acquire the ability to replicate in these cells, thus expanding their host-range (Ohlund et al., 2019). As such, a thorough reconstruction of evolutionary history could help in the identification of possible host switch over time, and how the virus will spread to new hosts. Additionally, phylodynamic analyses of genomic sequence data could be used to infer the time and location of their most recent common ancestors, their routes of dissemination, and demographic dynamics. However, these analyses require a wide range of temporal and geographic sampling of informative viral sequences.

Our genetic analyses of cISF were based on the assembly of multiple datasets of sequences, all of which grouped either *ns5* or ORF-coding sequences sharing common ancestry. The obtained results have shown that regardless of their overall high phylogenetic signal, a strong temporal signal could only be confirmed for two species-specific sequence datasets (ds1 and ds5, grouping CTFV and CFAV, respectively). Even though both of these corresponded to partial *ns5* genomic sequence datasets, one might expect that the use of longer genome regions might offer more phylogenetic resolution and increase the temporal signal (Pinel-Galzi et al., 2009). However, in our datasets, simply increasing sequence length, while associated with an increase in phylogenetic signal, did not result in a corresponding increase in temporal signal. We also observed that *ns5* and ORF-coding sequence-based tree topologies were always congruent. Flavivirus *ns5* sequences have been extensively used for phylogenetic and phylogeographic analyses, especially in datasets with incomplete genome sequences (Cella et al., 2019; Iwashita et al., 2018). Therefore, in this specific case, there seems to exist little to no advantages to the use of whole-genomic sequences when against *ns5* sequences, which are far more accessible.

Despite its caveats, the present study represents the first spatiotemporal dispersal of specific sublineages within the cISF radiation. The best candidate models for a Bayesian phylogeographic were determined, using the two analyzed datasets of viral sequences associated with a single host species. This is critical in such an analysis, as it has been proved that not only weak temporal signal may be reflected in uncertainty of date and/or rate estimation in a Bayesian coalescent framework, but coalescent priors may also affect mean tMRCA estimates (Trovao et al., 2015).

Whereas the spatiotemporal analyses of both CTFV and CFAV *ns5* sequence seemed to suggest expansion through distinct routes, the proposed ages for their respective mean MRCA may have been biased. The inner nodes of the MCC trees displayed large 95% HPD intervals for root ages, especially in the case of the CFAV dataset. Not surprisingly, when a similar approach was used to reconstruct the spatiotemporal diffusion of cISF associated with *Culex pipiens*, for which a root-to-tip analysis had indicated a low temporal signal, the combination of a recent root and a large age root 95% HPD interval was even more apparent.

The analysis of CTFV *ns5* sequences did place a tree root from which two genetic sublineages expanded into Southeast Asia and southern Europe. However, how the virus spread from its suggested shared ancestor into Europe remained unclear. Interestingly, the European *Cx. theileri* cISF seemed to have reached the Iberian Peninsula and Turkey on two separate occasions, but the geographic origin for their MRCA was not strongly supported. The time frame in which the latter events occurred was not clear either as estimated by the broad age root 95% HPD intervals. Regardless of the exact date, these viruses started to diverge from their MRCA, our analyses suggested that it may have occurred recently. These recent movements of mosquito-restricted viruses may benefit from climate changes, massive tourism, continuous population growth especially in urban areas, as well as global commercial trading which have also been blamed for the repeated introduction of *Aedes albopictus* into Europe, and its expansion in recent times (Parreira and Sousa, 2015). Moreover, while the obtained results support the formerly suggested dispersal of CTFV within the Iberian Peninsula, the Turkish origin for the Myanmar clade, as suggested previously (Cella et al., 2019), could not be confirmed.

Although CFAV was initially discovered in 1975 and appears to be ubiquitous in *Aedes aegypti* mosquitoes collected worldwide (Baidaliuk et al., 2019), very little is known about its evolutionary history. Our analysis placed the MRCA of all sequences analyzed in the 1930s, although, as mentioned above, this date may be biased, and is likely to be older. We estimated its expansion towards the Americas and Asia, and from there into Oceania or Africa. A similar dissemination pattern was suggested for CPFV, starting in Asia in the early 20th century (but as early as the late 1700s) and accelerating over the last two decades, and dispersing on a global scale. Considered altogether, our spatiotemporal analysis demonstrated multiple points of origin for cISF and diverse pathways of worldwide dissemination. These long-distance movements are unlikely to be solely carried out by insects (that have shorter migration routes), but most likely also benefit from human-associated activities and population movements.

Genetic diversity analyses were performed for both *ns5* partial genomic and whole-genome sequence datasets. Although the considered cut-off value was arbitrarily defined (minimum of 84% nucleotide sequence identity), based on the divergence estimates obtained, it seems suitable to define independently evolving genetic lineages. While analyzing *ns5* and complete genome sequences of *Culex*-specific cISF (and all cISF), little to no distinct patterns were found after all analyses are considered. Our analysis point that *ns5* was by far the genomic region with the lowest overall mean genetic diversity, which, interestingly, does not match results found by based on the analysis of CTFV (Bittar et al., 2016), whereas the *ns2a*-coding sequence was the region with the highest overall mean genetic distance. As expected, mean genetic distance values were consistently higher for coding regions of structural proteins when comparing them to non-structural proteins. Altogether, the cISF sequences were characterized by low (Shanon) entropy, displaying low variation in the amino acid sequences of their products, which probably results from low numbers of non-synonymous substitutions.

PCOORD analyses were executed side by side with phylogenetic reconstructions and they uncovered subtle patterns not found through conventional analyses, as was the case with CFAV. Sequences of the same sublineage usually display similar distribution patterns by PCOORD analysis, allowing the verification of how close different sublineages are to one another, which is especially useful for the analysis of nucleotide sequence datasets with low genetic variability. With CFAV, even though all sequences

appeared to be closely related through a phylogenetic reconstruction with a ML tree, PCOORD was able to unveil different patterns even between all CFAV.

Selective pressure analyses indicated that most of cISF genome is under strong purifying selection. Extensive research has shown over time that flavivirus can inhibit immune responses in their respective hosts to allow persisting infections, generating an advantageous immunological balance between the mosquito and the virus (Mukherjee et al., 2019). While this might suggest that this would lead to low selective pressure and a diversification of the viral population (Coffey et al., 2013), other studies have suggested that purifying selection is a major driver of arbovirus evolution (Lequime et al., 2016) even in invertebrate cells (Vasilakis et al., 2009). While cISF are not subject to the pressures of a natural maintenance cycle where they replicate in both vertebrates and invertebrates as arboviruses do, and may, therefore, evolve under distinct selective scenarios, our results suggest that a strong selective pressure does impact the evolution of cISF. However, the *ns2a* region seems to present higher ω values then other regions of the viral genome. Past studies suggest that flavivirus *ns2a* participates in the assembly and egress of the virion (Xie et al., 2019; Zhang et al., 2019), but while mutations on the *ns2a* protein impair virion morphogenesis, they have little to no effect on viral RNA synthesis (Xie et al., 2013). Additionally, reversion mutations in the NS3 coding sequence (one of the regions with the lowest ω values) could restore infectious virus production in virus harboring *ns2a* mutations (Liu et al., 2003). This could explain the more relaxed pressure against amino acid change in the *ns2a* region when compared to other regions.

This report did end up highlighting caveats affecting continuous phylogeography analyses when applied to cISF. Most datasets analysed did not show significant temporal signals, especially those including either all cISF or *Culex*/*Aedes*-specific ISF. While uncertainties in the estimates rates of nucleotide sequence substitution or tMRCA may result from weak temporal signals (Trovao et al., 2015), they may also be affected by the rapid rates of evolution of RNA viral genomes due to saturation. However, a negative impact in MRCA date estimates due to strong purifying selection seems to be negligible in the case of cISF. Besides, while the number of cISF sequences has been growing in recent years, host-associated, geographic, and temporal biases (with most sequences being obtained quite recently), most probably impact these analyses. Regardless of its

limitations, the information generated in this study could help define how these types of analyses may be conducted in future studies and highlight the need for a correct assessment of the best candidate models for Bayesian time-calibrated phylogenetic and phylogeographic analyses, as different candidate models may apply to different sequence datasets.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work received financial support from the Global Health and Tropical Medicine Center, which is funded through FCT contract UID/ Multi/04413/2013. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

CRediT authorship contribution statement

Paulo Morais: data curation and analysis, visualization and writing of the manuscript; Nídia S. Trovão and Ana Barroso Abecasis: revision and edition of the manuscript; Ricardo Parreira: conceptualization, supervision, edition and revision of the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2021.198507.

References

Abílio, A.P., Silva, M., Kampango, A., Narciso, I., Gudo, E.S., Das Neves, L.C.B., Sidat, M., Fafetine, J.M., De Almeida, A.P.G., Parreira, R., 2020. A survey of RNA viruses in mosquitoes from Mozambique reveals novel genetic lineages of flaviviruses and phenuiviruses, as well as frequent flavivirus-like viral DNA forms in Mansonia. BMC Microbiol 20, 1–15. https://doi.org/10.1186/s12866-020-01905-5.

Araújo, J.M.G., Nogueira, R.M.R., Schatzmayr, H.G., Zanotto, P.M.d.A., Bello, G., 2009. Phylogeography and evolutionary history of dengue virus type 3. Infect. Genet. Evol. 9, 716–725. https://doi.org/10.1016/j.meegid.2008.10.005.

Baidaliuk, A., Miot, E.F., Lequime, S., Moltini-Conclois, I., Delaigue, F., Dabo, S., Dickson, L.B., Aubry, F., Merkling, S.H., Cao-Lormeau, V.-M., Lambrechtsa, L., 2019. Cell-fusing agent virus reduces arbovirus dissemination in Aedes aegypti mosquitoes. In Vivo, 93, 1-17.

Barzilai, L.P., Schrago, C.G., 2019a. The range of sampling times affects Zika virus evolutionary rates and divergence times. Arch. Virol. 164, 3027–3034. https://doi.org/10.1007/s00705-019-04430-7.

Barzilai, L.P., Schrago, C.G., 2019b. The range of sampling times affects Zika virus evolutionary rates and divergence times. Arch. Virol. 164, 3027–3034. https://doi.org/10.1007/s00705-019-04430-7.

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., Myers, M.F., George, D.B., Jaenisch, T., William Wint, G.R., Simmons, C.P., Scott, T.W., Farrar, J.J., Hay, S.I., 2013. The global distribution and burden of dengue. Nature 496, 504–507. https://doi.org/10.1038/nature12060.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2016. SpreaD3: interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Mol. Biol. Evol. 33, 2167–2169. https://doi.org/10.1093/molbev/msw082.

Bittar, C., Machado, D.C., Vedovello, D., Ullmann, L.S., Rahal, P., Araújo Junior, J.P., Nogueira, M.L., 2016. Genome sequencing and genetic characterization of Culex Flavirirus (CxFV) provides new information about its genotypes. Virol. J. 13 https://doi.org/10.1186/s12985-016-0614-3.

Blitvich, B., Firth, A., 2015. Insect-specific flaviviruses: a systematic review of theirm discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. Viruses. https://doi.org/10.3390/v7041927.

Calzolari, M., Ze-Ze, L., Vazquez, A., Sanchez Seco, M.P., Amaro, F., Dottori, M., 2016. Insect-specific flaviviruses, a worldwide widespread group of viruses only detected in insects. Infect. Genet. Evol. 40, 381–388. https://doi.org/10.1016/j.meegid.2015.07.032.

Castresana, Jose, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Cella, E., Benvenuto, D., Donati, D., Garilli, F., Angeletti, S., Pascarella, S., Ciccozzi, M., 2019. Phylogeny of culex theileri virus flavivirus in Spain, Myanmar, Portugal and Turkey. Asian Pac. J. Trop. Med. 12, 216–223.

Chong, H.Y., Leow, C.Y., Abdul Majeed, A.B., Leow, C.H., 2019. Flavivirus infection—a review of immunopathogenesis, immunological response, and immunodiagnosis. Virus Res. 274, 197770. https://doi.org/10.1016/j.virusres.2019.197770.

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

Colmant, A.M.G., Hobson-Peters, J., Bielefeldt-Ohmann, H., van den Hurk, A.F., Hall- Mendelin, S., Chow, W.K., Johansen, C.A., Fros, J., Simmonds, P., Watterson, D., Cazier, C., Etebari, K., Asgari, S., Schulz, B.L., Beebe, N., Vet, L.J., Piyasena, T.B., Nguyen, H.D., Barnard, R.T., Hall, R.A., 2017. A new clade of insect-specific flaviviruses from Australian anopheles mosquitoes displays species-specific host restriction. mSphere 2, e00262-17. https://doi:10.1128/mSphere.00262-17.

Cook, S., Moureau, G., Kitchen, A., Gould, E.A., de Lamballerie, X., Holmes, E.C., Harbach, R.E., 2012. Molecular evolution of the insect-specific flaviviruses. J. Gen. Virol. 93, 223–234. https://doi.org/10.1099/vir.0.036525-0.

Crabtree, M.B., Sang, R.C., Stollar, V., Dunster, L.M., Miller, B.R., 2003. Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. Arch. Virol. 148, 1095–1118. https://doi.org/10.1007/s00705-003-0019-7.

Crochu, S., Cook, S., Attoui, H., Charrel, R.N., De Chesse, R., Belhouchet, M., Lemasson, J.J., de Micco, P., de Lamballerie, X., 2004. Sequences of flavivirus related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J. Gen. Virol. 85, 1971–1980. https://doi.org/10.1099/vir.0.79850-0.

Drummond, a.J., Rambaut, a., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192. https://doi.org/10.1093/molbev/msi103.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006a. Relaxed phylogenetics and dating with confidence. PLoS Biol 4, 699–710. https://doi.org/10.1371/journal.pbio.0040088.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006b. Relaxed phylogenetics and dating with confidence. PLoS Biol 4, 699–710. https://doi.org/10.1371/journal.pbio.0040088.

Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. Trends Ecol. Evol. 18, 481–488. https://doi.org/10.1016/S0169-5347(03)00216-7.

Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., Suchard, M.A., 2013. Improving bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724. https://doi.org/10.1093/molbev/mss265.

Goenaga, S., Goenaga, J., Boaglio, E.R., Enria, D.A., Levis, S.del C., 2020. Superinfection exclusion studies using West Nile virus and Culex flavivirus strains from Argentina. Mem. Inst. Oswaldo Cruz 115, 1–5. https://doi.org/10.1590/0074-02760200012.

Griffiths, R.C., Tavare, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 344, 403–410. https://doi.org/10.1098/rstb.1994.0079.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Harrisom, J., Hobson-Peters, J., Colmant, A., Koh, J., Newton, N., Warrilo, D., Bielefeldt- Ohmann, H., Piyasena, T., O'Brien, C., Vet, L., Paramitha, D., Potter, J., Davis, S., Johansen, C., Setoh, Y., Khromykh, A., 2020. Antigenic characterization of new lineage ii insect-specific flaviviruses in australian mosquitoes and identification of host restriction factors. mSphere 5, 1–19.

Ho, S.Y.W., Phillips, M.J., Drummond, A.J., Cooper, A., 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Mol. Biol. Evol. 22, 1355–1363. https://doi.org/10.1093/molbev/msi125.

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., Melville, L., Hunt, N., Webb, R.I., Blitvich, B.J., Whelan, P., Hall, R.A., 2013. A new insect-specific flavivirus from northern Australia suppresses replication of west Nile virus and Murray valley encephalitis virus in Coinfected mosquito cells. PLoS One 8, 1–12. https://doi.org/10.1371/journal.pone.0056534.

Holbrook, M.R., 2017. Historical perspectives on flavivirus research. Viruses 9, 1–19. https://doi.org/10.3390/v9050097.

Iwashita, H., Higa, Y., Futami, K., Lutiali, P.A., Njenga, S.M., Nabeshima, T., Minakawa, N., 2018. Mosquito arbovirus survey in selected areas of Kenya: Detection of insect-specific virus. Trop. Med. Health 46, 1–15. https://doi.org/10.1186/s41182-018-0095-8.

Junglen, S., Korries, M., Grasse, W., Wieseler, J., Kopp, A., Hermanns, K., Leon-Juarez, M., Drosten, C., Kümmerer, B.M., 2017. Host range restriction of insect specific flaviviruses occurs at several levels of the viral life cycle. mSphere 2, 1–15. https://doi.org/10.1128/msphere.00375-16.

Katoh, K., Standley, D.M., 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.

Kosakovsky Pond, S.L., Frost, S.D.W., 2005. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533. https://doi.org/10.1093/bioinformatics/bti320.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Kuno, G., Chang, G.J., Tsuchiya, K.R., Karabatsos, N., Cropp, C.B., 1998. Phylogeny of the genus Flavivirus. J. Virol. 72, 73–83.

Kuwata, R., Isawa, H., Hoshino, K., Sasaki, T., Kobayashi, M., Maeda, K., Sawabe, K., 2015. Analysis of mosquito-borne flavivirus superinfection in culex tritaeniorhynchus (Diptera: Culicidae) cells persistently infected with culex flavivirus (Flaviviridae). J. Med. Entomol. 52, 222–229. https://doi.org/10.1093/jme/tju059.

Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk in continuous space and time. Mol. Biol. Evol. 27, 1877–1885. https://doi.org/10.1093/molbev/msq067.

Lequime, S., Fontaine, A., Ar Gouilh, M., Moltini-Conclois, I., Lambrechts, L., 2016. Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. PLoS Genet 12, e1006111. https://doi.org/10.1371/journal.pgen.1006111.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1, 1–5. https://doi.org/10.1093/ve/vev003.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471. https://doi.org/10.1093/molbev/msn090.

Mukherjee, D., Das, S., Begum, F., Mal, S., Ray, U., 2019. The Mosquito Immune System and the Life of Dengue Virus: What We Know and do not Know. Pathogens. https://doi.org/10.3390/pathogens80200778.

Nei, M., Gojoborit, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410.

Nouri, S., Matsumura, E.E., Kuo, Y.W., Falk, B.W., 2018. Insect-specific viruses: from discovery to potential translational applications. Curr. Opin. Virol. 33, 33–41. https://doi.org/10.1016/j.coviro.2018.07.006.

Ohlund, P., Lunden, H., Blomstrom, A.L., 2019. Insect-specific virus evolution and potential effects on vector competence. Virus Genes 55, 127–137. https://doi.org/10.1007/s11262-018-01629-9.

Parreira, R., Sousa, C.A., 2015. Dengue fever in Europe: could there be an epidemic in the future? Expert Rev. Anti. Infect. Ther. 13, 29–40. https://doi.org/10.1586/14787210.2015.982094.

Peterson, A.T., 2014. Defining viral species: making taxonomy useful. Virol. J. 11, 1–4. https://doi.org/10.1186/1743-422X-11-131.

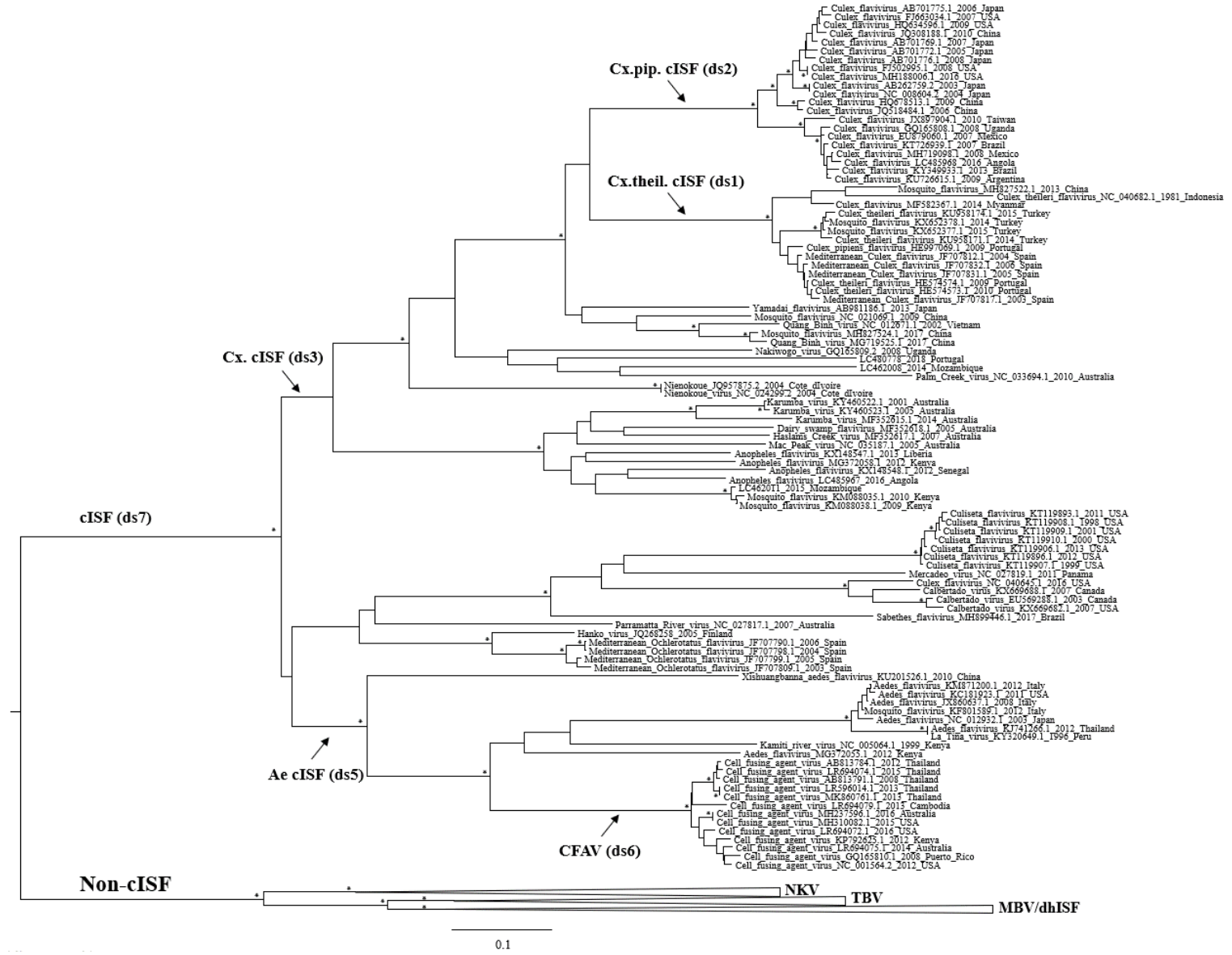
Pinel-Galzi, A., Mpunami, A., Sangu, E., Rakotomalala, M., Traore, O., Sereme, D., Sorho, F., Sere, Y., Kanyeka, Z., Konate, G., Fargette, D., 2009. Recombination, selection and clock-like evolution of Rice yellow mottle virus. Virology 394,164–172. https://doi.org/10.1016/j.virol.2009.08.008.

Rambaut, A., Lam, T.T., Carvalho, L.M., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2, 1–7. https://doi.org/10.1093/ve/vew007.

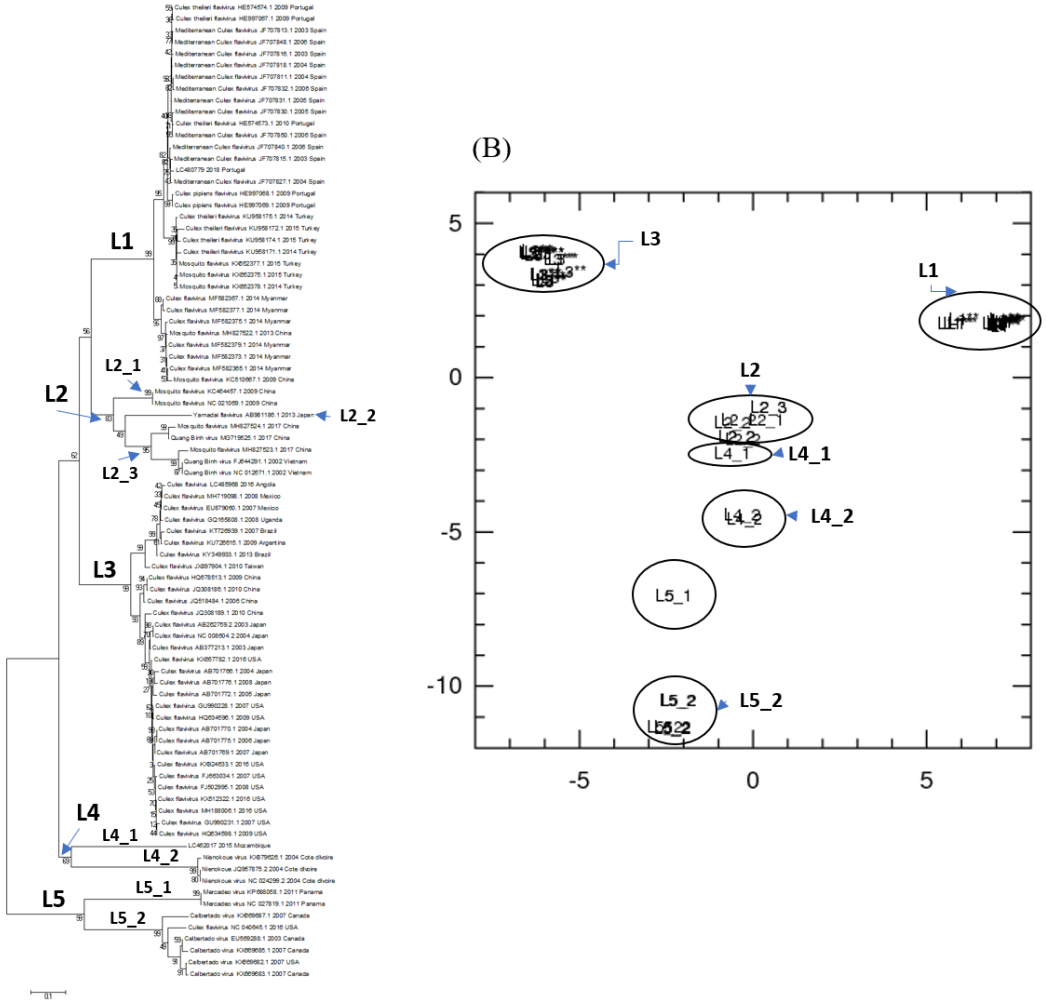
Romo, H., Kenney, J.L., Blitvich, B.J., Brault, A.C., 2018. Restriction of Zika virus infection and transmission in Aedes aegypti mediated by an insect-specific flavivirus. Emerg. Microbes Infect. 7. https://doi:10.1038/s41426-018-0180-4.

Sall, A.A., Faye, O., Diallo, M., Firth, C., Kitchen, A., Holmes, E.C., 2010. Yellow fever virus exhibits slower evolutionary dynamics than Dengue Virus. J. Virol. 84,765–772. https://doi.org/10.1128/jvi.01738-09.

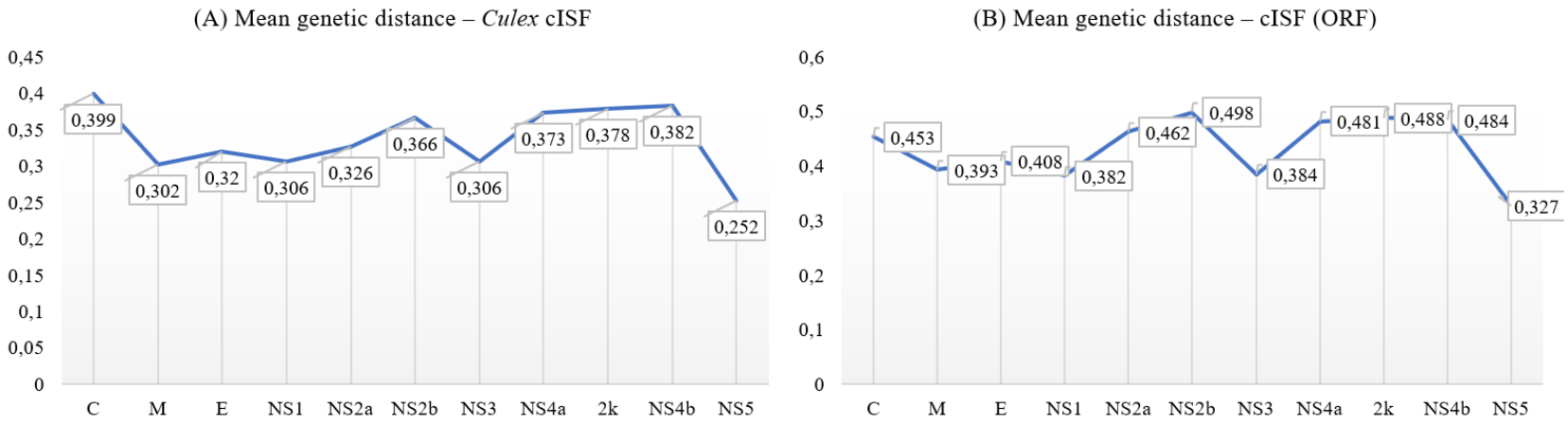
Supplementary Fig. 1: Maximum likelihood tree of flaviviruses' ns5 coding sequences. At specific branches, an asterisk (*) indicates ≥75% of bootstrap support. The branches defining the datasets of ns5 sequences analysed are also indicated: *Culex theileri* flavivirus *NS5* sequences (Cx.theil cISF; ds1), *Culex pipiens* flavivirus *NS5* sequences (Cx.pip. cISF; ds2), *Culex* flavivirus *NS5* sequences (Cx. cISF; ds3), *Aedes* flavivirus *NS5* sequences (Ae. cISF; ds5), cell fusing agent virus *NS5* sequences (CFAV; ds6), cISF ns5 sequences including a maximum of two sequences from the same genetic lineage per country per year of sampling (cISF; ds7). In the tree, the non-cISF radiation includes mosquito-borne viruses [**MBV**: Sepik virus (NC_008719), Wesselsbron virus (NC_012735), Yellow fever virus (NC_002031), West Nile virus (NC_009942), West Nile virus (NC_001563), Koutango virus (EU082200), Yaounde virus (NC_034018), Japanese encephalitis virus (NC_001437), Usutu virus (NC_006551), Murray Valley encephalitis virus virus (NC_000943), Alfuy virus (AY898809), Cacipacore virus virus (NC_026623), Saint Louis encephalitis virus (NC_007580), Dengue virus 1 (NC_001477), Dengue virus 2 (NC_001474), Dengue virus 3 (NC_001475), Dengue virus 4 (NC_002640), Zika virus (NC_012532), Spondweni virus (NC_029055), Tembusu virus (NC_015843), Ntaya virus (NC_018705), Bagaza virus (NC_012534), THo virus (EU879061), Rocio virus (AY632542), Ilheus virus (NC_009028), Iguape virus (AY632538), Naranjal virus (KF917538), Bussuquara virus (NC_009026), Kedougou virus (NC_012533), New Mapoon virus (NC_032088), Sokoluk bat virus (NC_026624), Entebbe bat virus (NC_008718), Yokose virus (NC_005039), Stratford virus (KM225263), Torres virus (KM225265), Kokobera virus (NC_009029), Bainyik virus (KM225264), Potiskum virus (NC_029054), Saboya virus (NC_033697), Jugra virus (NC_033699), Uganda S virus (NC_033698), Bamaga virus (NC_033725), Edge Hill virus (NC_030289), Paraiso Escondido virus (NC_027999)], tick-borne viruses [**TBV**: Saumarez Reef virus (DQ235150), Kama virus (NC_023439), Tyuleniy virus (NC_023424), Kadam virus (NC_033724), Meaban virus (NC_033721), Powassan virus (NC_003687), Gadgets Gully virus (NC_033723), Royal Farm virus (DQ235149), Langat virus (NC_003690), Turkish sheep encephalitis virus (DQ235151), Tick-borne encephalitis virus (NC_001672), Spanish goat encephalitis virus (NC_027709), Louping ill virus (NC_001809)], dual-host affiliated insect specific viruses [**dhISF**: Nhumirim virus (NC_024017), Kampung Karu virus (KY320648), Barkedji virus (KC496020), Nanay virus (MF139575), Nounane virus (NC_033715), Chaoyang virus (FJ883471), Chaoyang virus (NC_017086), Long Pine Key virus (KY290254), Lammi virus (NC_024806), Donggang virus (NC_016997), Marisma mosquito virus (MF139576), Ilomantsi virus (KC692067)] and no-known vector viruses [**NKV**: Apoi virus (NC_003676), Rio Bravo virus (NC_003675), Phnom Penh bat virus (NC_034007), Modoc virus (NC_003635), Montana myotis leukoencephalitis virus (NC_004119), Jutiapa virus (NC_026620)].


Culex_flavivirus_AB701775.1_2006_Japan
Culex_flavivirus_FJ663034.1_2007_USA
Culex_flavivirus_HQ634596.1_2009_USA
Culex_flavivirus_JQ308188.1_2010_China
Culex_flavivirus_AB701769.1_2007_Japan
Culex_flavivirus_AB701772.1_2005_Japan
Culex_flavivirus_AB701776.1_2008_Japan
Culex_flavivirus_FJ502995.1_2008_USA
Culex_flavivirus_MH188006.1_2018_USA
Culex_flavivirus_AB262759.2_2009_Japan
Culex_flavivirus_NC_008604.2_2004_Japan
Culex_flavivirus_HQ678513.1_2009_China
Culex_flavivirus_JQ518484.1_2006_China
Culex_flavivirus_JX887904.1_2010_Taiwan
Culex_flavivirus_GQ165808.1_2008_Uganda
Culex_flavivirus_EU879060.1_2007_Mexico
Culex_flavivirus_KT726939.1_2007_Brazil
Culex_flavivirus_MH719098.1_2008_Mexico
Culex_flavivirus_LC485968_2016_Angola
Culex_flavivirus_KY349933.1_2013_Brazil
Culex_flavivirus_KU726615.1_2009_Argentina
Cx.pip. cISF (ds2)
Mosquito_flavivirus_MH827522.1_2013_China
Culex_theileri_flavivirus_NC_040682.1_1981_Indonesia
Culex_flavivirus_MF582367.1_2014_Myanmar
Culex_theileri_flavivirus_KU958174.1_2015_Turkey
Mosquito_flavivirus_KX652378.1_2014_Turkey
Mosquito_flavivirus_KX652377.1_2015_Turkey
Culex_theileri_flavivirus_KU958171.1_2014_Turkey
Culex_pipiens_flavivirus_HE997069.1_2009_Portugal
Mediterranean_Culex_flavivirus_JF707812.1_2004_Spain
Mediterranean_Culex_flavivirus_JF707832.1_2005_Spain
Mediterranean_Culex_flavivirus_JF707831.1_2005_Spain
Culex_theileri_flavivirus_HE574574.1_2009_Portugal
Culex_theileri_flavivirus_HE574573.1_2010_Portugal
Mediterranean_Culex_flavivirus_JF707817.1_2005_Spain
Cx.theil. cISF (ds1)
Yamadai_flavivirus_AB981186.1_2013_Japan
Mosquito_flavivirus_NC_021069.1_2009_China
Quang_Binh_virus_NC_012671.1_2002_Vietnam
Mosquito_flavivirus_MH827524.1_2017_China
Quang_Binh_virus_MG719525.1_2017_China
Nakiwogo_virus_GQ165809.2_2008_Uganda
LC480778_2018_Portugal
LC462008_2014_Mozambique
Palm_Creek_virus_NC_033694.1_2010_Australia
Nienokoue_JQ957875.2_2004_Cote_dIvoire
Nienokoue_virus_NC_024299.2_2004_Cote_dIvoire
Karumba_virus_KY460522.1_2001_Australia
Karumba_virus_KY460523.1_2003_Australia
Karumba_virus_MF552615.1_2014_Australia
Dairy_swamp_flavivirus_MF552618.1_2005_Australia
Haslams_Creek_virus_MF552617.1_2007_Australia
Mac_Peak_virus_NC_035187.1_2005_Australia
Anopheles_flavivirus_KX148547.1_2013_Liberia
Anopheles_flavivirus_MG372058.1_2012_Kenya
Anopheles_flavivirus_KX148548.1_2012_Senegal
Anopheles_flavivirus_LC485967_2016_Angola
LC462071_2013_Mozambique
Mosquito_flavivirus_KM088035.1_2010_Kenya
Mosquito_flavivirus_KM088038.1_2009_Kenya
Cx. cISF (ds3)
cISF (ds7)
Culiseta_flavivirus_KT119893.1_2011_USA
Culiseta_flavivirus_KT119908.1_1998_USA
Culiseta_flavivirus_KT119909.1_2001_USA
Culiseta_flavivirus_KT119910.1_2000_USA
Culiseta_flavivirus_KT119906.1_2013_USA
Culiseta_flavivirus_KT119896.1_2012_USA
Culiseta_flavivirus_KT119907.1_1999_USA
Mercadeo_virus_NC_027819.1_2011_Panama
Culex_flavivirus_NC_040645.1_2016_USA
Calbertado_virus_KX669688.1_2007_Canada
Calbertado_virus_EU569288.1_2003_Canada
Calbertado_virus_KX669682.1_2007_USA
Sabethes_flavivirus_MH899446.1_2017_Brazil
Parramatta_River_virus_NC_027817.1_2007_Australia
Hanko_virus_JQ268258_2005_Finland
Mediterranean_Ochlerotatus_flavivirus_JF707790.1_2006_Spain
Mediterranean_Ochlerotatus_flavivirus_JF707798.1_2004_Spain
Mediterranean_Ochlerotatus_flavivirus_JF707799.1_2005_Spain
Mediterranean_Ochlerotatus_flavivirus_JF707809.1_2003_Spain
Xishuangbanna_aedes_flavivirus_KU201526.1_2010_China
Aedes_flavivirus_KM871200.1_2012_Italy
Aedes_flavivirus_KC181923.1_2011_USA
Aedes_flavivirus_JX860637.1_2008_Italy
Mosquito_flavivirus_KF801589.1_2012_Italy
Aedes_flavivirus_NC_012932.1_2005_Japan
Aedes_flavivirus_KJ741266.1_2012_Thailand
La_Tina_virus_KY320649.1_1996_Peru
Kamiti_river_virus_NC_005064.1_1999_Kenya
Aedes_flavivirus_MG372053.1_2012_Kenya
Cell_fusing_agent_virus_AB813784.1_2012_Thailand
Cell_fusing_agent_virus_LR694074.1_2015_Thailand
Cell_fusing_agent_virus_AB813791.1_2008_Thailand
Cell_fusing_agent_virus_LR596014.1_2013_Thailand
Cell_fusing_agent_virus_MK860761.1_2015_Thailand
Cell_fusing_agent_virus_LR694079.1_2015_Cambodia
Cell_fusing_agent_virus_MH237596.1_2016_Australia
Cell_fusing_agent_virus_MH310082.1_2015_USA
Cell_fusing_agent_virus_LR694072.1_2016_USA
Cell_fusing_agent_virus_KP792625.1_2012_Kenya
Cell_fusing_agent_virus_LR694075.1_2014_Australia
Cell_fusing_agent_virus_GQ165810.1_2008_Puerto_Rico
Cell_fusing_agent_virus_NC_001564.2_2012_USA
Ae cISF (ds5)
CFAV (ds6)
Non-cISF
NKV
TBV
MBV/dhISF
0.1

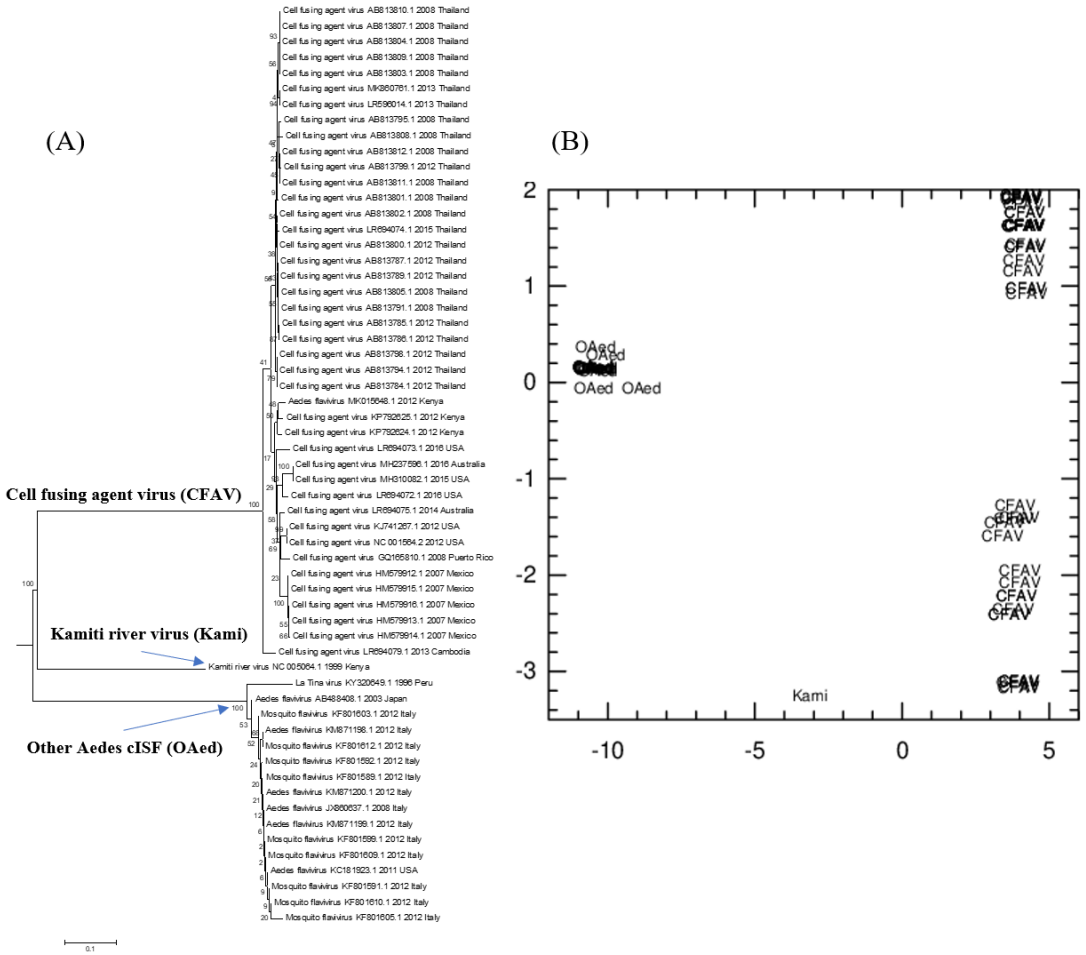
Supplementary Fig. 2: (A) Maximum likelihood tree for *Culex* cISF (partial ns5 coding sequence). The different sublineages are identified as L1-L5; (B) Principal coordinate analysis carried out for partial NS5 *Culex* cISF coding sequences using PCOORD. Each sequence is identified by the sublineage they belong to (as shown in Supplementary Figure 2A).



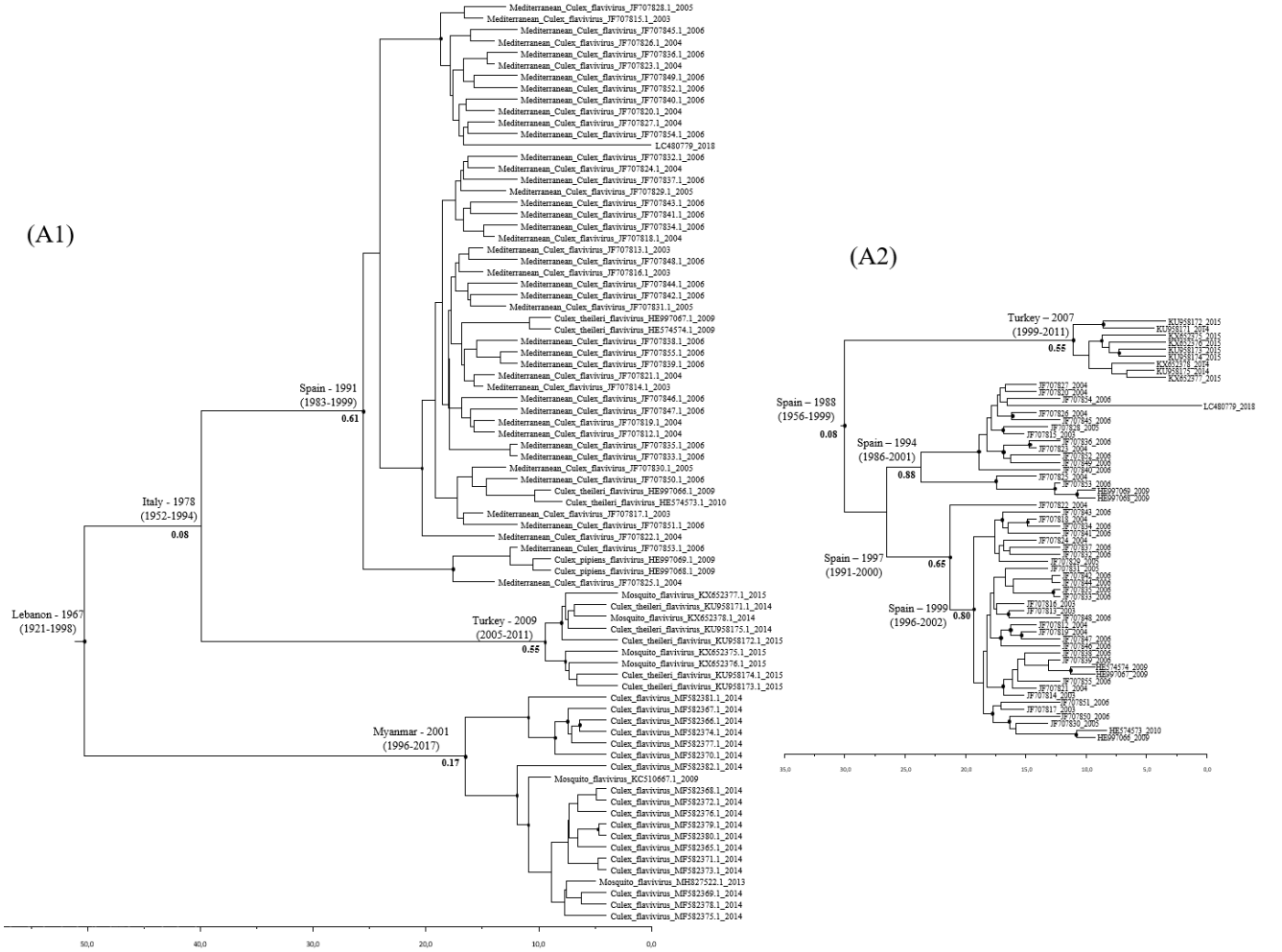
Supplementary Fig. 3: Analysis of the mean genetic distance values (K2-P corrected) of the different coding sections of *Culex*-specific (A) and *Culex/Aedes/Anopheles* cISF (B), based on the analysis of alignments of nucleotide sequences.



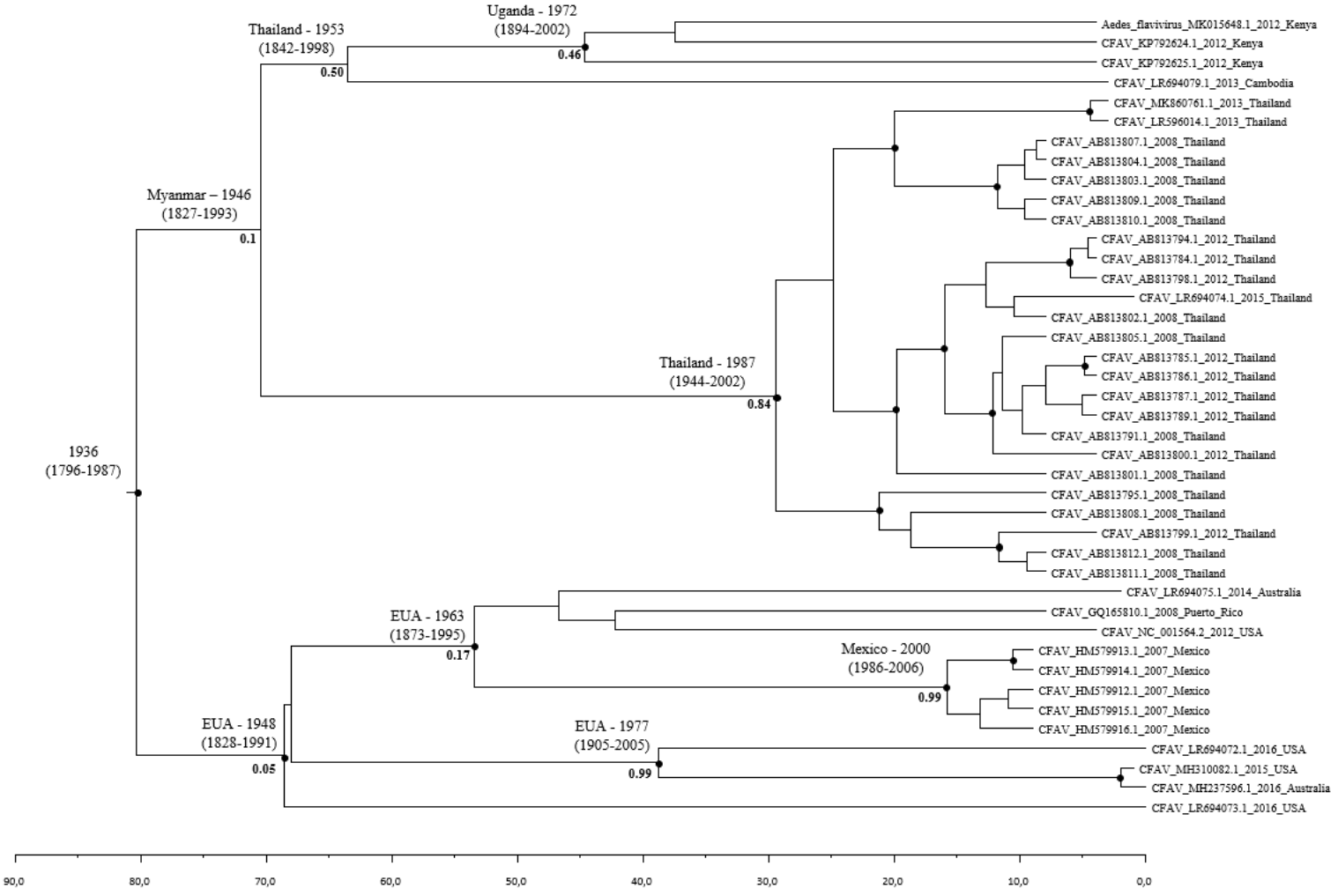
Supplementary Fig. 4: (A) Maximum likelihood tree for Aedes cISF (partial NS5 coding sequence). The different sublineages are identified as CFAV (Cell Fusing Agent Virus), Kami (Kamiti River Virus), and OAed (other Aedes cISF); (B) Principal coordinate analysis carried out for partial NS5 Aedes cISF coding sequences using PCOORD. Each sequence is identified by the sublineage they belong to (as shown in Supplementary Figure 3A).



Supplementary Fig. 5: Continuous phylogeography tree representing the geographic spread of *Culex theileri* cISF (dataset ds1, corresponding to partial NS5 coding sequence). At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles, while the decimals associated with certain nodes indicate the PP for the suggested location. A1 and A2 refer to the analysis of the entire CTFV dataset, or only those sequences with a European origin, respectively.

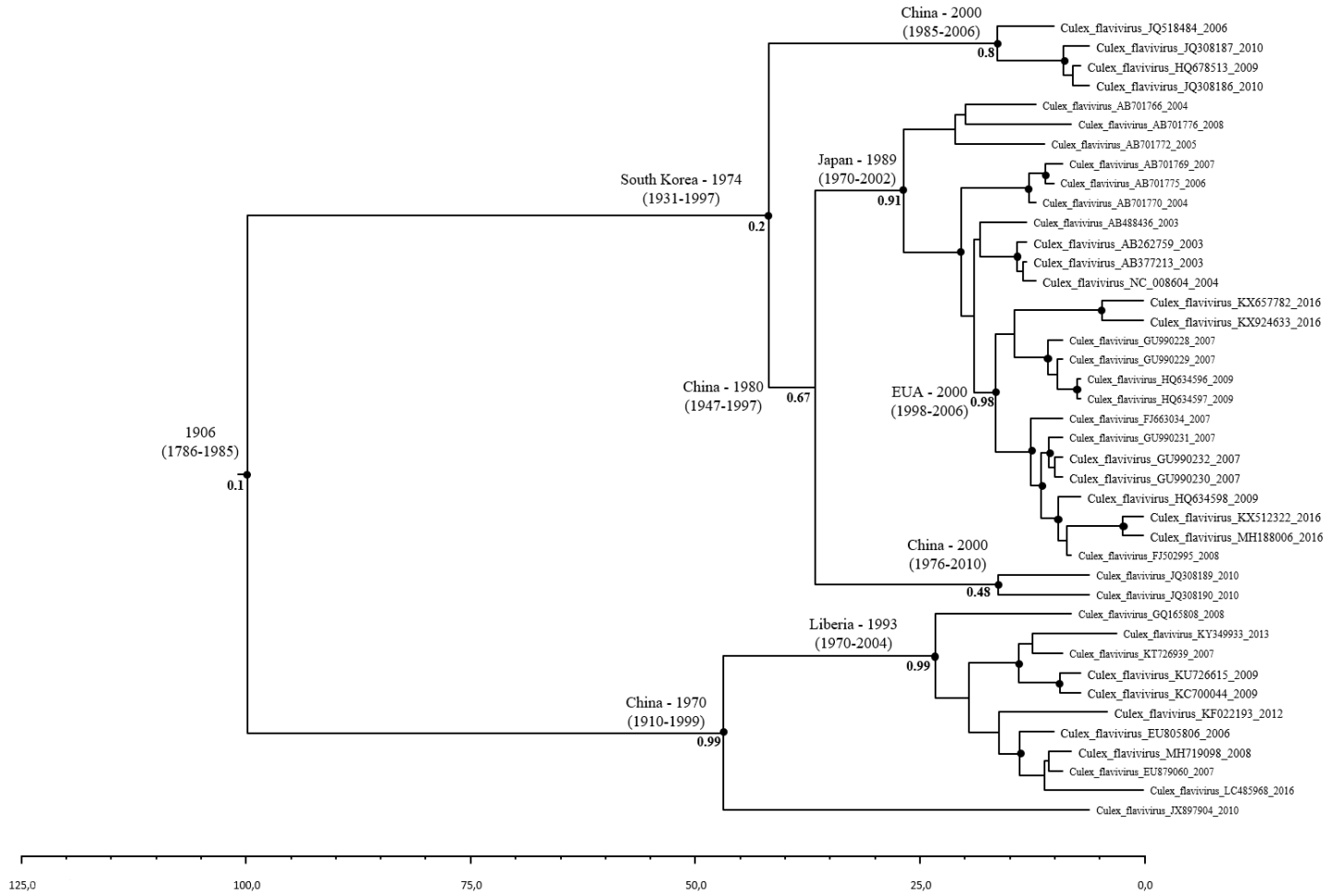


Supplementary Fig. 6: Continuous phylogeographic analysis of partial CFAV *ns5* coding sequence (corresponding to dataset ds5). At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles, while the decimals associated with certain branches indicate the PP for the suggested location.



Supplementary Fig. 7: (A) Continuous phylogeographic analysis of partial CPFV *ns5* coding sequence (corresponding to dataset ds2). At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles, while the decimals associated with certain branches indicate the PP for the suggested location. (B) Spatiotemporal reconstruction of CPFV spread visualized on SpreaD3 software using the MCC tree represented in Figure 7A.

(A)



Supplementary Table 1: Assessment of selective pressure of cISF by three different methods (Single Likelihood Ancestor Counting. SLAC. and Fixed Effects Likelihood. FEL. in the DataMonkey server; and by the Synonymous Non-synonymous Analysis Program. SNAP. by the HIV Los Alamos Database) of full genome and each genomic region. with a p value of 0.05.

GO	Complete genome	C	M	E	NS1	NS2a	NS2b	NS3	NS4a	2k	NS4b	NS5
FEL	0.039	0.138	0.048	0.035	0.027	0.382	0.113	0.024	0.058	0.028	0.049	0.029
SLAC	0.110	0.223	0.133	0.117	0.101	0.453	0.218	0.091	0.161	0.151	0.131	0.076
SNAP	0.118	0.369	0.253	0.235	0.133	0.813	0.514	0.13	0.368	0.684	0.32	0.024
Sites under negative selection	**Complete genome**	**C**	**M**	**E**	**NS1**	**NS2a**	**NS2b**	**NS3**	**NS4a**	**2k**	**NS4b**	**NS5**
Number of sites	**2774**	**38**	**118**	**400**	**303**	**122**	**114**	**507**	**61**	**23**	**197**	**876**
FEL	2554	32	107	386	290	51	87	493	51	21	187	811
SLAC	2254	19	90	328	268	27	53	452	37	19	165	755
Sites under positive selection	**Complete genome**	**C**	**M**	**E**	**NS1**	**NS2a**	**NS2b**	**NS3**	**NS4a**	**2k**	**NS4b**	**NS5**
Number of sites	**2774**	**38**	**118**	**400**	**303**	**122**	**114**	**507**	**61**	**23**	**197**	**876**
FEL	21	0	0	0	0	14	0	0	1	0	1	0
SLAC	4	0	0	0	0	9	0	0	2	0	0	0

Supplementary Table 2: a) Assessment of selective pressure of *Culex* cISF by three different methods (Single Likelihood Ancestor Counting. SLAC. and Fixed Effects Likelihood. FEL. in the DataMonkey server; and by the Synonymous Non-synonymous Analysis Program. SNAP. by the HIV Los Alamos Database) of full genome and each genomic region. with a p value of 0.05; b) Assessment of selective pressure of *Aedes* cISF and CFAV by three different methods (Single Likelihood Ancestor Counting. SLAC. and Fixed Effects Likelihood. FEL. in the DataMonkey server; and by the Synonymous Non-synonymous Analysis Program. SNAP. by the HIV Los Alamos Database). Partial ns5 coding sequence. 241 codons (Aedes cISF) and 258 codons (CFAV) analyzed with a p value of 0.05.

a)

GO	Complete genome	C	M	E	NS1	NS2a	NS2b	NS3	NS4a	2k	NS4b	NS5
FEL	0.044	0.292	0.038	0.034	0.027	0.177	0.140	0.030	0.085	0.036	0.059	0.026
SLAC	0.117	0.358	0.107	0.119	0.104	0.242	0.233	0.102	0.173	0.126	0.141	0.075
SNAP	0.067	0.394	0.155	0.226	0.142	0.308	0.444	0.128	0.328	0.761	0.252	0.020
Sites under negative selection	**Complete genome**	**C**	**M**	**E**	**NS1**	**NS2a**	**NS2b**	**NS3**	**NS4a**	**2k**	**NS4b**	**NS5**
Number of sites	**3121**	**38**	**134**	**421**	**366**	**176**	**111**	**571**	**110**	**23**	**239**	**887**
FEL	2290	16	105	336	289	63	45	466	73	16	173	722
SLAC	1580	8	71	202	200	31	21	326	43	9	103	583
Sites under positive selection	**Complete genome**	**C**	**M**	**E**	**NS1**	**NS2a**	**NS2b**	**NS3**	**NS4a**	**2k**	**NS4b**	**NS5**
Number of sites	**3121**	**38**	**134**	**421**	**366**	**176**	**111**	**571**	**110**	**23**	**239**	**887**
FEL	10	0	0	2	0	2	1	0	1	0	0	0
SLAC	0	0	0	0	0	0	0	0	0	0	0	0

b)

ω	FEL	SLAC	SNAP
CFAV	0.097	0.105	0.046
Aedes	0.038	0.073	0.042
Sites under negative selection	**FEL**	**SLAC**	
CFAV	36 (13.9%)	11 (4.3%)	
Aedes	130 (53.9%)	49 (20.3%)	
Sites under positive selection	**FEL**	**SLAC**	
CFAV	0	0	
Aedes	0	0	

Supplementary Table 3: Assessment of phylogenetic signal of a) *Culex*-specific cISF (ds7) and b) cISF (ds8) sequences by likelihood mapping analysis using whole-genome (ORF-coding) and protein coding segments (C to NS5).

a)	**Datasets**											
***Culex* cISF (ds7)**	**ORF**	**C**	**M**	**E**	**NS1**	**NS2a**	**NS2b**	**NS3**	**NS4a**	**2k**	**NS4b**	**NS5**
Totally resolved quartets	99.6%	86.4%	89.8%	96.6%	97%	88.3%	89.8%	97.8%	89.6%	81.8%	95.3%	99%
Partially resolved quartets	0.3%	4.8%	4.9%	1.7%	1.7%	5.7%	3.6%	0.9%	3.1%	4.6%	2.2%	0.8%
Unresolved quartets	0.1%	8.9%	5.3%	1.7%	1.3%	6.0%	6.6%	1.4%	7.3%	13.7%	2.5%	0.3%
b)	**Datasets**											
cISF (ds8)	**ORF**	**C**	**M**	**E**	**NS1**	**NS2a**	**NS2b**	**NS3**	**NS4a**	**2k**	**NS4b**	**NS5**
Totally resolved quartets	99.6%	80.6%	90.4%	98%	97.7%	89%	89.4%	98.6%	83.8%	83.5%	95%	99.3%
Partially resolved quartets	0.3%	7.4%	4.1%	1.0%	1.4%	5%	5.3%	0.8%	6.8%	8.6%	2.6%	0.7%
Unresolved quartets	0.0%	12.1%	5.5%	0.9%	0.9%	5.8%	5.3%	0.7%	9.5%	7.9%	2.5%	0.1%

ORF: viral polyprotein; C: capsid protein; M: membrane glycoprotein; E: envelope protein; NS: non-structural; 2k: 2kD peptide.

Supplementary Table 4: Performance of differences coalescent priors and geographic diffusion priors: (a) Comparative analysis of the performance of parametric and non-parametric coalescent priors in the analysis of cISF sequences (ds1, ds2 and ds5); (b) Evaluation of rates for coalescent combined with different geographic diffusion priors: analysis of CTFV (ds1) and CFAV (ds5) *ns5* sequences.

(a)

	root_age [95% HPD]	mean rate [95% HPD]	stdev [95% HPD]	coefficient of variation [95% HPD]	PS1	PS2	SS1	SS2
ds1								
Constant	1936 [1874–1983]	8.97E-04 [5.25E-04–1.27E-03]	1.33E-03 [4.71E-04–2.51E-03]	1.03 [0.63–1.46]	-3988.96	-3989.09	-3989.21	-3989.33
Exponential	1963 [1930–1987]	9.88E-04 [6.38E-04–1.35E-03]	1.21E-03 [4.34E-04–2.23E-03]	0.97 [0.62–1.37]	-3994.41	-3990.07	-3994.71	-3990.30
Expansion	1938 [1880–1984]	8.99E-04 [5.37E-04–1.28E-03]	1.26E-03 [4.09E-04–2.36E-03]	1.00 [0.62–1.42]	-3989.69	-3989.01	-3989.92	-3989.23
GMRF Skyride	1992 [1986–1997]	1.23E-03 [8.21E-04–1.62E-03]	2.37E-03 [9.45E-04–4.37E-03]	1.45 [0.99–1.96]	-3984.75	-3985.31	-3985.17	-3985.64
Skygrid	1947 [1891–1987]	9.59E-04 [5.69E-04–1.42E-03]	1.37E-03 [4.49E-04–2.76E-03]	1.03 [0.63–1.47]	-3989.76	-3991.99	-3989.89	-3992.36
Skyline	1966 [1926–1998]	1.13E-03 [6.45E-04–1.67E-03]	1.45E-03 [4.34E-04–2.98E-03]	1.03 [0.62–1.51]	-3983.76	-3979.52	-3979.94	-3984.05
ds2								
Constant	1779 [1556–1959]	4.66E-04 [9.67E-05–8.46E-04]	2.34E-04 [4.55E-09–6.94E-04]	0.39 [1.9E-05–0.87]	-2229.64	-2229.44	-2229.66	-2229.48
Exponential	1869 [1728–1999]	5.79E-04 [1.45E-04–1.07E-03]	9.21E-04 [1.09E-08–1.23E-03]	0.45 [3.74E-05–1.08]	-2233.42	-2233.47	-2233.54	-2233.57
Expansion	1814	4.64-04	2.28E-04	0.37	-2231.14	-2229.32	-2231.38	-2229.33

	[1606–1954]	[1.25E-04–8.18E-04]	[4.71E-10–6.65E-04]	[7.1E-06–0.86]				
GMRF Skyride	1969 [1928–1998]	1.07E-03 [2.50E-04–1.93E-03]	2.34E-04 [4.55E-09–6.94E-04]	1.54 [0.59–2.67]	-2238.04	-2236.73	-2238.04	-2236.78
Skygrid	1717 [1558–1965]	4.62E-04 [9.73E-05–8.49E-04]	9.21E-04 [1.09E-08–1.23E-03]	0.37 [2.02E-05–0.85]	-1923.83	-2158.01	-1923.84	-2158.09
Skyline	1698 [1261–1984]	4.07E-04 [1.62E-05–8.05E-04]	2.28E-04 [4.71E-10–6.65E-04]	0.38 [1.94E-06–0.89]	-2232.44	-2231.58	-2232.52	-2231.72
ds5								
GMRF Skyride	1865 [1767–1984]	3.24E-04 [5.80E-05–6.27E-04]	2.40E-07 [4.9E-324–1.3E-55]	1.54E-03 [0–1.94E-15]	-2521,8	-2520,32	-2521,91	-2520,39
Skygrid	-4.67E32 [-3.64E-32–1978]	9.03E-05 [1.19E-38–3.51E-04]	8.26E-08 [4.9E-324–7.60E-41]	1.57E-03 [0–1.99E-15]	-1789,79	-1702,87	-1789,81	-1702,87
Skyline	-4.97E17 [-3.39E17–1079]	7.12E-05 [1.42E-23–2.71E-04]	6.49E-07 [4.9E-324–1.16E-12]	9.62E-03 [0–4.41E-05]	-2517,83	-2518,25	-2517,95	-2518,38

(b)

	root_age [95% HPD]	**mean rate [95% HPD]**	**stdev [95% HPD]**	**coefficient of variation [95% HPD]**	**PS1**	**PS2**	**SS1**	**SS2**
ds1								
Constant+ Brownian	1195 [680–1629]	2.66E-04 [1.22E-04–3.87E-04]	5.29E-03 [1.50E-03–0.01]	1.78 [1.21–2.45]	-4230.33	-4231.45	-4228.12	-4230.33
Constant+ CauchyRRW	1936 [1875–1983]	9.62E-04 [5.72E-04–1.33E-03]	1.53E-03 [5.50E-04–2.84E-03]	1.08 [0.70–1.50]	-4097.57	-4094.1	-4094.57	-4093.46
GMRF Skyride+	1918	1.17E-03	1.09E-02	1.98	-4319.57	-4324.48	-4318.21	-4323.85

Brownian	[1861–1965]	[6.79E-04–1.67E-03]	[1.84E-03–0.03]	[1.22–2.82]				
GMRF Skyride+ CauchyRRW	1993 [1987–1998]	1.30E-03 [8.89E-04–1.68E-03]	2.65E-03 [1.04E-03–4.89E-03]	1.49 [1.03–2.02]	-4090.37	-4089.63	-4089.77	-4089.02
Skyline+ Brownian	748 [-3360–1391]	1.86E-04 [8.49E-05–2.61E-04]	8.73E-03 [2.87E-03–0.02]	1.96 [1.35–2.62]	-4213.98	-4208.93	-4212.45	-4208.33
Skyline+ CauchyRRW	1964 [1923–1998]	1.15E-03 [6.81E-04–1.64E-03]	1.58E-03 [5.28E-04–3.14E-03]	1.07 [0.69–1.51]	-4081.08	-4094.97	-4080.92	-4093.67
ds5								
GMRF Skyride+ CauchyRRW	-65696 [1796–1987]	3.8E-04 [6.82E-05–6.94E-04]	2.68E-05 [4.9E-324–1.67E-04]	0.0538 [0–0.46]	-2681.21	-2685.44	-2680.82	-2685.02
GMRF Skyride+ GammaRRW	-4.43E8 [1588–1999]	3.38E-04 [6.17E-15–6.14E-04]	2.05E-05 [4.9E-324–1.31E-04]	0.0451 [0–0.43]	-2674.85	nd	-2674.99	nd
GMRF Skyride+ LognormalRRW	782 [1820–1986]	4.12E-04 [9.01E-5–7.45E-04]	1.61E-05 [4.9E-324–8.92E-05]	0.028 [0–0.28]	-2739.45	-2743.72	-2738.93	-2743.24
GMRF Skyride+ Brownian	-6.32E-14 [-1.6E-12–1987]	4.74E-05 [1.93E-20–8.42E-05]	1.27E-03 [1.48E-19–4.45E-03]	2.06 [1.20–3.08]	-2842.71	-2844.65	-2842.33	-2844.17

GMRF: gaussian Markov random field; HPD: highest Probability Density; nd: not determined; PS: path sampling; RRW: Relaxed Random Walk; SS: stepping-stone sampling. The values 1 and 2 associated with PS and SS indicate those obtained in two independent MCMC runs.

Chapter 5. Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Published as:

Morais, P., Trovão, N. S., Abecasis, A. B., & Parreira, R. (2022). Readdressing the genetic diversity and taxonomy of the Mesoniviridae family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences. Virus Research, 313 (March), 198727. https://doi.org/10.1016/j.virusres.2022.198727

Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Abstract

Research on the recently established *Mesoniviridae* family (Order *Nidovirales*), RNA genome insect-specific viruses, has been steadily growing in the last decade. However, after the last detailed phylogenetic characterization of mesoniviruses in 2014, numerous new sequences, even in organisms other than mosquitos, have been identified and characterized.

In this study, we analyzed nucleotide and protein sequences of mesoniviruses with a wide range of molecular tools including genetic distance, Shannon entropy, selective pressure analysis, polymorphism identification, principal coordinate analysis, likelihood mapping and phylodynamic reconstruction. We also sought to reevaluate new mesoniviruses sequence positions within the family, proposing a taxonomic revision.

The different sub-lineages of mosquito mesoniviruses sequences presented low sequence diversity and entropy, with incongruences to the existing taxonomy being found after an extensive phylogenetic characterization. High sequence discrepancy and differences in genome organization were found between mosquito mesoniviruses and other mesoniviruses, so their future classification, as other meso-like viruses that are found in other organisms, should be approached with caution.

No evidence of frequent recombination was found, and mesonivirus genomes seem to evolve under strong purifying selection. Insufficient data by root-to-tip analysis did not yet allow for an adequate phylogeographic reconstruction.

Keywords: Mesonivirus; Nidovirales; Taxonomy; Genetic diversity; Phylogenetic analysis

1. Introduction

The Order *Nidovirales* comprises a genetically diverse assemblage of enveloped, approximately spherical viruses with linear single-stranded, positive-sense, and polyadenylated RNA genomes, that can infect a wide range of hosts, from mammals to insects. According to the International Committee on Taxonomy of Viruses (ICTV), they are taxonomically (mid-2021) distributed in eight suborders and 14 families (https://talk.ictvonline.org/taxonomy/), including the well-studied *Arteriviridae, Coronaviridae*, and *Roniviridae,* as well as the more recently established *Mesoniviridae* family (Vasilakis et al., 2014).

Within the Order *Nidovirales*, mesoniviruses were the first known to infect insects. Their detailed description was initiated in 2011 with the characterization of the Cavally (CAVV) and Nam Dinh (NDiV) viruses, isolated from *Culex* mosquitos, collected in Cote d'Ivoire and Vietnam, respectively (Zirkel et al., 2011; Nga et al., 2011). Since then, mesoniviruses have been isolated from mosquitos collected in the Americas (Charles et al., 2018), Asia (Wang et al., 2017), Africa (Diagne et al., 2020), and Oceania (Warrilow et al., 2014), suggesting a global distribution. Like insect-specific flaviviruses (Blitvich and Firth, 2015) and mosquito-associated bunyaviruses (Marklewitz et al., 2013), mesoniviruses are considered some of the most predominant RNA genome insect-specific viruses (ISVs) (Vasilakis et al., 2014). While they have repeatedly been isolated from naturally infected mosquitoes, they do not appear to infect vertebrates (Blitvich and Firth, 2015). Nonetheless, their isolation from *Aphis citricidus* aphids collected in 2012 in China suggests that the host range of mesoniviruses might go beyond that which is currently known (Chang et al., 2020). Furthermore, a meso-like virus has already been detected in Italy in *Leveillula taurica,* an obligate fungal pathogen (accession number MN609866).

The genomes of mesoniviruses of approximately 20 kb are organized into multiple open-reading frames (ORFs). The most frequently found organization is ORF1a-ORF1b-ORF2a-ORF2b-ORF3a-ORF3b-ORF4, but exceptions do exist (e.g., the Meno virus does not encode ORF4). The 5′ region of the genome encodes two polyproteins (ORF1a and ORF1b), the expression of which is controlled by ribosomal frameshift followed by proteolytic processing (Vasilakis et al., 2014), and their products are suggested to be involved in the regulation of gene expression, polyprotein processing, and genome

replication and transcription. The 3′ region of the genome includes smaller ORFs that encode structural proteins. Apart from ORF1a and ORF1b, the number of small ORFs varies among different viruses in the Order *Nidovirales* (Gorbalenya et al., 2006).

The latest update from the ICTV regarding the *Mesoniviridae* family (March 2021, available at https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/308/mesoniviridae), acknowledges 1 single genus (*Alphamesonivirus*) and 8 subgenera. *Namcalivirus* is represented by both the Alphamesonivirus 1 species (comprising most mesoniviruses isolated to date), and the Alphamesonivirus 10 species (which includes the Dianke virus). Other genera encompass only one other viral type. For example, the Ofaie virus (OFAV) and the Casuarina virus (CASV) are currently the sole representatives of the *Ofalivirus* (Alphamesonivirus 6) and *Casualivirus* (Alphamesonivirus 4) genera, respectively. In addition, several recently discovered mesoniviruses [e.g., the Odorna virus (OdoV)], remain unclassified.

Considering the recent pandemic spread of SARS-CoV-2 coronavirus the interest in the study of mesoniviruses has increased (Lai et al., 2020). While distantly related to coronaviruses and mostly restricted to mosquitoes, their study might hold crucial information regarding the evolution of the viruses within the Order *Nidovirales*, as they may have evolved in arthropods (Nga et al., 2011). However, while the genomic and phylogenetic characterization of mesonivirus has lastly been addressed in a comparative dating from 2014 (Vasilakis et al., 2014), since then, the isolation of multiple mesoniviruses prompted us to reevaluate their position within the family. Furthermore, the recent discovery of a meso-like virus in organisms other than mosquitos might hold new information regarding their phylogenetic relationship with other mesoniviruses. In this report, we will also discuss the conditions required for a potential future phylogeographic analyses of this *taxa*.

2. Materials and methods

2.1. Dataset preparation and sequence alignments

The compilation of the different nucleotide (nt) and amino acid (aa) sequence datasets used in this work was based on the selection of complete genome sequences available at GenBank in 01/07/2021. These were either directly identified via their respective accession numbers, or indirectly singled out as a product of similarity searches using BLASTn.

All sequences corresponding to complete genomes available were downloaded, and additional information including GenBank accession number, host species, geographic origin, and collection date was obtained. When available, information on genomic coding-capacity (ORF organization) and their respective sequences were also collected. Furthermore, for comparative purposes, representative datasets containing ORF1ab nt and aa sequences of viruses from other families in the Order *Nidovirales* (corresponding to the most conservative coding region between them) were also constructed.

Multiple alignments of complete nt and aa sequences were performed using the iterative G-INS-I method as implemented in MAFFT v.7 (Katoh and Standley, 2013), followed by their edition using GBlocks (Castresana, 2000), allowing for less strict flanking positions in the obtained multiple sequence alignments (MSA). These were systematically verified to ensure the correct alignment of homologous codons using BioEdit 7.0.5 (Hall, 1999). Additional alignments were also constructed for different ORFs identified in the *Mesoniviridae* family that included ORF1a, ORF1b, ORF2a/spike, ORF3a, and ORF4, as well as the viral RNA-dependent RNA polymerase (RdRp). Multiple alignments of ORF1ab aa sequences from different families in the *Nidovirales* Order were performed similarly using MAFFT iterative L-INS-I option, followed by a new alignment using the G-INS-I method.

2.2. Assessment of the temporal and phylogenetic signals of different mesonivirus sequence datasets

The evolutionary information contained in all used sequence datasets (phylogenetic signal) was assessed by Likelihood Mapping (Strimmer and von Haeseler, 1997) using

TREE-PUZZLE v5.3 (Schmidt et al., 2002). Datasets with totally resolved quartets values of over 90% were considered of high phylogenetic resolving power.

A visual inspection of the degree of temporal signal (i.e., signal for divergence accumulation over the sampling time interval) in the complete genome nt datasets (as well as for the RdRp and S protein-coding sequences) for all mesoniviruses was carried out using an exploratory linear regression approach assuming the topology of a Maximum Likelihood (ML) tree, estimated under a non-clock (unconstrained) and the GTR+Γ+I substitution model using IQ-TREE (Trifinopoulos et al., 2016). Root-to-tip divergences were plotted as a function of sampling time using the TempEst v. 1.5.3 (Rambaut et al., 2016).

2.3. Phylogenetic analyses using maximum likelihood

Phylogenetic reconstructions of full-length genomic nt and ORF-specific nt datasets and specific aa sequences (RdRp and S datasets) were performed based on the maximum likelihood optimization criterion, using the GTR+Γ+I model and Whelan And Goldman (WAG) model, respectively, as suggested by IQ-TREE (Trifinopoulos et al., 2016), which was also used for tree building. The stability of the obtained tree topologies was assessed by bootstrapping and using the aLRT (approximate likelihood ratio test) with 1000 re-samplings of the original sequence data.

2.4. Genetic diversity and protein primary sequence analyses

The estimation of genetic distance values (corrected with the Kimura-2P formula) was carried out using MEGAX (Kumar et al., 2018). Heat maps were calculated based on pairwise evolutionary distances obtained using the Heatmapper webserver (Babicki et al., 2016), while box plots were drawn with Microsoft Excel. Visualization of viral genomic organization was performed using Open Reading Frame Finder (available in https://www.ncbi.nlm.nih.gov/orffinder/), while the SMART webserver (Letunic and Bork, 2018) was used for the identification and annotation of genetically mobile domains. The presence of conserved domains in viral protein sequences was investigated using CD-Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Remote homology

detection and structure prediction was analyzed using HHblits and Hpred, as well as sensitive sequence searching by HHMER (Zimmermann et al., 2018). Several bioinformatic tools were employed to investigate ORFs with unknown function, including computation of molecular weight and theoretical isoelectric point (pI) via ProtParam (https://web.expasy.org/protparam/), analysis of hydropathicity by ProtScale (https://web.expasy.org/cgi-bin/protsc ale/protscale.pl), prediction of transmembrane helices via TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM), prediction of n-glycosylation sites (http://www.cbs.dtu.dk/services/NetNGlyc) and o-glycosylation sites (http://www.cbs.dtu.dk/services/NetOGlyc), signal sequence search by SignalP (http://www.cbs.dtu.dk/service s/SignalP) and protein subcellular localization prediction by DeepLoc (http://www.cbs.dtu.dk/services/DeepLoc). The detection of repeats in protein sequences was carried out with RADAR (https://www.ebi.ac. uk/Tools/pfa/radar/).

The analyses of selective pressure on individual sites of codon alignments were carried out using the Single Likelihood Ancestor Counting (SLAC), the Fixed Effects Likelihood methods as implemented in Datamonkey (Kosakovsky Pond and Frost, 2005), or the SNAP tool (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html) that explores a simple method for calculation of synonymous and non-synonymous substitutions (Nei and Gojoborit, 1986). The degree of variability at each amino acid position in multiple alignments of single ORF aa sequences was evaluated based on the Shannon entropy function using Entropy (Shannon entropy-one option, available at http://www. hiv.lanl.gov/content/sequence/ENTROPY/entropy.html). Finally, principal coordinate analysis was carried out using PCOORD (http://www. hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html). Possible recombination events were investigated using the Recombination Detection Program 4 (RDP4) software (Martin et al., 2015).

2.5. Comparative analysis with virus from other Nidovirales families

For comparative analyses of mesonivirus genomic sequences with those of other nidoviruses, overall mean distances, assessment of phylogenetic signals, and selective

pressure analyses were performed for families in the *Nidovirales* Order with higher representation in the genomic sequence databases (*Arteriviridae, Coronaviridae,* and *Tobaniviridae*), focusing on the most conserved coding region among them (ORF1ab).

To assess the relationship between different families in the *Nidovirales* Order, phylogenetic reconstructions were carried out using multiple sequence alignments of RdRp aa sequences as described in Section 2.3.

3. Results

3.1. Comparative genome organization analyses

A total of 47 full-length mesonivirus genomic sequences, downloaded from the public genomic databases, were aligned, and analyzed. These included both those that, until 2020, had been only identified in multiple species of mosquitoes (n = 44), being frequently associated with *Culex sp.* or *Aedes sp.* In addition, this dataset also included three meso-like viral sequences that had recently been identified in hosts other than mosquitoes. These comprised those of meso-like viruses isolated from *Aphis citricidus* aphids (*Aphis citricidus* meson-like virus, AcMSV), from *Thrips tabaci* thrips (Insect metagenomics mesonivirus 1, Immeso1; Chiapello et al., 2021), as well as from a fungal pathogen, *Leveillula Taurica* (*Leveillula taurica* associated alphamesonivirus 1, temporarily abbreviated as LtM). All these have been listed in Supplementary Table 1. Additionally, for phylogenetic and other comparative analyses, ORF1ab aa sequences were also compiled for viruses in other families in the *Nidovirales* Order, and these have been included in Supplementary Table 2. Alongside the full-length genome datasets, other datasets including the nt and aa sequences of all mesonivirus identifiable ORFs (of the sequences listed in Supplementary Table 1) were also constructed.

Also, as suggested by Gorbalenya et al. in 2006, and as corroborated here in Supplementary Fig. 1, the number of ORFs identified in viral genomes from viruses allocated to different families in the *Nidovirales* Order is substantially different. Viruses from the *Mesoniviridae* family display smaller genomes with 4 to 7 ORFs, as similarly observed in the *Tobaniviridae* and *Medoniviridae* families. In contrast, viruses from the *Arteriviridae* and *Coronaviridae* families have a larger number of ORFs, up to 12.

A comparative analysis of the organization predicted for the different mosquito mesoniviruses (MM) genomes (Supplementary Fig. 1) indicated overall conserved synteny, with only those of meno-, kadiweu- and ofaieviruses missing an identifiable ORF4. As no complete genomic sequence have yet been made available for OdoV, a prediction of its genomic structures remains incomplete. A comparison between the genome organization of MM and other mesoniviruses (OM) revealed substantial differences, especially considering their similar genome sizes (excluding the 3'-poly [A] tail, they range from 19,209 nt for Immeso1 to 20,626 nt for AcMSV). All these viruses are suggested to use ribosomal frameshifting for translational control of the expression of non-structural proteins, as revealed by the consistent overlap between ORF1a and ORF1b, while ORF2a (surface spike) encodes the S glycoprotein. As expected, and considering that the mature products of ORF1a and 1b are usually involved in the control of essential steps of the viral replication cycle that include genomic replication, transcription, RNA capping and polyprotein processing, the genomic organization appears similar when MM and OM are compared (although smaller in size in LtM), including most conserved domains and other so-called genetically mobile domains (i.e., transposable elements; Vasilakis et al., 2014) (Fig. 1). Only Immeso1 does not seem to possess a coiled coil motif in ORF1a, while displaying a zinc finger domain which, however, is not shared by other mesoniviruses. Unfortunately, the available LtM genomic sequence appears to be incomplete, with only the full sequence of ORF1a and a partial sequence of ORF1b currently available. While the ORFs at the 3' half of the genome of MM were similar (except for ORF4), the number and organization of the ORFs identified in that same region of the genomes of OM are different. These ORFs (identified as ORFs x1-3 and ORFs y1-3 in Fig. 1) seem to encode putative products that, in most cases, share no easily identifiable homology with other viral proteins, nor do they display readily recognizable conserved domains (as defined by CD-BLAST analysis) associated with a particular biological function or protein family. Sequence searches regarding both y1 and y3 did not return any putative matches with homologues in the sequence databases, not even when remote homology detection methods were used (HHblits, HHpred, or HMMER). However, y1 is predicted as a 22kDa, basic (pI 9.9) and hydrophilic, while y3 is also small (12kDa) and basic but mostly hydrophobic. Furthermore, while multiple O-glycosylation sites were predicted in y1, none have been predicted for y3. On the other

hand, y2 is larger (98 kDa) and mostly neutral. However, remote homology detection tools indicated a 96% probability match between the highly basic (pI=10.5) product of ORFx2 (between amino acids 108 and 216) and the putative nucleocapsid protein of the Kadiweu virus (Alphamesonivirus 7), while part the product of ORFy2 aligned with the ORF2a protein encoded by NDiV (positions 391 and 906). In addition, the larger ORF found in the genomes of the *Aphis citricidus* meson-like virus and the *Thrips tabaci* associated mesonivirus (ORFs x3 and y2 in Fig. 1) encode putative proteins with 3 (y2) or 6 (x3) transmembrane helixes and multiple targets for N-glycosylation, which are features frequently found in integral viral envelope proteins. Finally, the putative ORF4, which is encoded by the genomes of mosquito mesoniviruses, is highly conserved (96.3% identity conservation among MM). It encodes a small (approximately 5kDa), basic (pI=9.6) hydrophilic protein, with no glycosylated amino acids, transmembrane helices, or signal peptide sequences (the latter found in the product of *Aphis citricidus* meson-like virus ORFx1), or conserved domains. Up to the present, its function remains unknown. In addition, insertions blocks in ORF1a have been described in a handful of mesonivirus (Kamphaeng Phet, KPhV; Karang Sari virus, KSaV; Bontang Baru, BBaV; Vasilakis et al., 2014), but among the more recently identified MM and OM, the Dak Nong virus (DkNV) also revealed those same type of insertions. Two of these insertion blocks are larger (approximately 570 nt) than the other (approximately 170 nt) but despite their size difference they partially align at their 5′ ends. All these insertions extend the coding capacity of ORF1a, with the larger of these two insertions, found in the genomes of the BBaV and KSaV viruses, apparently coding for hydrophilic peptides of approximately 190 aa characterized by two types of partially repeated sequences. One of these is repeated 3 to 4 times at the N-terminal section, while the other (in two copies), can be identified at the peptide's C-terminus. Homology searches did not reveal sequence similarities or amino acid motifs that might indicate their putative function, but the sequence encoded by the BBaV is characterized by the presence of a PKR13108 superfamily sequence motif found in prolipoprotein diacylglyceryl transferases, found in bacteria of the *Corynebacterineae* family ($E=7.64e^{-04}$).

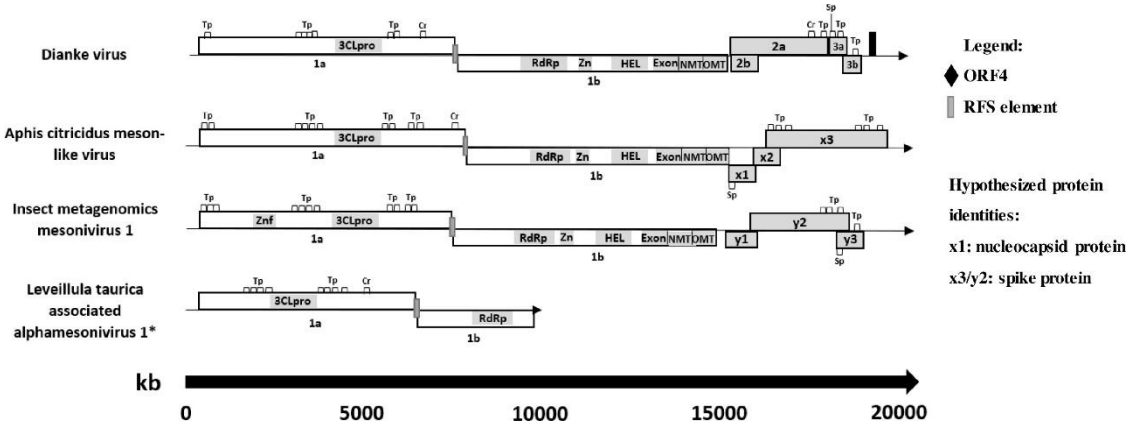


Fig. 1: Schematic representation of the genomic organization of mosquito mesoniviruses (Dianke virus, accession number MN622133, used as an example) and other mesoniviruses: Aphis citricidus meson-like virus (AcMSV, accession number MN961271); Insect metagenomics mesonivirus 1 (Immeso1, accession number MN714662); Leveillula taurica associated alphamesonivirus 1 (LtM, accession number MN609866); Znf = Zinc finger; 3CLpro = 3C-like protease; RdRp = RNA-dependent RNA polymerase; Zn = Zinc-binding domain; HEL = Helicase; Exon = 3′ -5′ exoribonuclease; NMT = N7-methyltransferase; OMT = Cap-0 specific (nucleoside-2′ -O-)-methyltransferase; Tp = Transmembrane region; Cr = Coiled coil region; Sp = Signal peptide. * - Leveillula taurica associated alphamesonivirus 1 whole genome sequence still not available. ORFs at the 3' half of the genome for other mesoniviruses identified as x1-x3 (for AcMSV) and y1-y3 (for Immeso1), with most recognizable putative proteins for each (when available) displayed.

3.2. Genetic diversity analyses

Overall mean genetic distances for MM were calculated for both the complete genome as well as for each ORF-specific genomic region (Supplementary Table 3). Furthermore, both genetic distances between all MM full-length genomes as well as for two ORF-specific genomic regions (RdRp and S) were also calculated (Supplementary Table 4). The overall mean distance (complete genome) between all sequences was 0.15. The region encompassing the ORF4 gene was the viral genomic region with the lowest mean genetic distance value (0.04), while the ORF1a region held the highest (0.17). Using ORF1ab aa sequences as a reference, analyses of datasets including up to three different sequences genus/subgenus selected from the viral families with a larger representation in the public databases (*Mesoniviridae, Coronaviridae, Arteriviridae* e *Tobaniviridae*), overall mean genetic distance values were 0.07, 0.34, 0.43 and 0.50, respectively. Pairwise evolutionary distances (PEDs) were calculated between all RdRp aa mesonivirus sequences (Supplementary Table 5), with heat maps designed to visualize intersequence genetic diversity, and box-and-whisker graphs used to visualize all

distances between mesoniviruses from the same species, between mesonivirus from different species and also between MM and OM. These analyses clearly highlight the difference between MM and OM, as seen in Supplementary Fig. 2. Substantial differences in PEDs between all groups analyzed were also highlighted using box-and-whisker graphs (Supplementary Fig. 3). While, and as expected, higher PEDs were obtained when protein sequence comparisons included viral sequences from different species (as opposed to intra virus species comparisons), with values always below the 96.8% protein sequence identity used as cut-off (Vasilakis et al., 2014), but this distance was far considerably pronounced when comparing MM to OM, with identity values under 70%. Two groups of mesoniviruses sequences, BBaV and KPhV, clearly exceed the cut-off value, with identity values below 80%, which suggests they should correspond to new species.

Shannon entropy is a quantitative measure of uncertainty in a dataset of nucleotide or amino acid sequences, and it may be considered as a measure of variation in DNA and protein sequence alignments for assessment of genetic diversity in a cross-sectional sense. When applied to the analysis of MM mesonivirus sequences, Shannon entropy calculations showed low values for all mesonivirus ORF-coding sequences. However, statistically higher entropy values were calculated for ORF1a, especially when compared to other genomic regions, while ORF3 showed the lowest entropy. Other families in the *Nidovirales* Order had consistently higher entropy values when compared with the *Mesoniviridae* family (Supplementary Fig. 4).

3.3. Phylogenetic analyses of mosquito mesoniviruses

For different datasets of nt mesonivirus sequences, likelihood mapping analyses were performed to calculate their respective phylogenetic signals (Table 1). The obtained results showed an overall high percentage (>90%) of the totally resolved sequence quartets (assessing the topologies of 10,000 quartet replicates) obtained for the complete genome, ORF1a, ORF1b and ORF2a, as well as the specific RdRp-coding sequence, while lower values were obtained for ORF3a (81.8%) and ORF4 (74.1%). These results indicate that most phylogenetic reconstructions based on the analysis of alignments for any viral ORFs other than ORF3 and 4 could be used to produce unambiguous trees.

Overall, phylogenetic reconstructions using full-length genome sequences from viruses allocated to different families in the *Nidovirales* Order consistently presented high phylogenetic signals (Supplementary Table 6). However, while the great majority of the constructed datasets revealed consistent high phylogenetic signals, standard linear regression exploration of root-to-tip distances as a function of sampling time (Supplementary Table 7) carried out to establish to what extent the *Mesoniviridae* family contained detectable signal for sequence divergence throughout the sampling time intervals, showed negative slope and correlation coefficient values, even after an extensive root-to-tip analysis and the removal of outlier sequences. This observation extended for both the full-genome, RdRp and S protein-coding sequence comparisons, as well as when analyzing only the Alphamesonivirus 1 species or all MM and OM at once (indicated as all mesoniviruses in Supplementary Table 7). As such, at the present, only the investigation of phylogenetic relationships using ML trees is possible, while potential temporal and phylogeographical analyses using a Bayesian phylodynamic framework await the description of future new mesonivirus sequences.

Table 1: Phylogenetic signal of mesonivirus sequence datasets.

	Datasets						
	Full-length genome	**ORF1a**	**ORF1b**	**RdRp**	**ORF2a (S)**	**ORF3a**	**ORF4**
Totally resolved quartets	99.3%	98.4%	98.7%	94,4%	96.8%	81.8%	74.1%
Partially resolved quartets	0.6%	0.8%	0.8%	2.7%	1.2%	3.7%	1.7%
Unresolved quartets	0.1%	0.9%	0.4%	2.9%	2.1%	14.4%	24.2%

Selective pressure analyses were carried by estimating omega (ω, i.e., dN/dS or the frequency of non-conservative-to-conservative substitutions ratio) values using concatenated ORF1a/ORF1b/ORF2/ ORF3/ORF4 coding-sequence datasets, as well as for each one of the individual ORFs using three different methods (SLAC, FEL and SNAP). No significant differences were found between all ORFs, apart from ORF4

(which seems to be the only region under diversifying selection, with ω values over 1, even though it is the genomic region with lowest genetic diversity), with all ω values being very low (Supplementary Table 8a), with site-specific selection analysis revealing high percentage of negatively selected sites, as well as very low percentage of positively selected sites. Comparative analyses with other families in the *Nidovirales* Order were performed using their most conservative region among their genomes, the ORF1ab region (Supplementary Table 8b). Unlike mesonivirus sequences, those from coronaviruses, tobaniviruses and arteriviruses displayed higher ω values, always higher than 1, with lower percentage of negatively selected sites and higher percentage of positively selected sites.

Recombination events are common among viruses classified into the *Nidovirales* Order (Gorbalenya et al., 2006) and has been shown to affect the evolution of some of its best studied members (Hon et al., 2008; Boni et al., 2020). Since no previous assessment of whether these events affect the evolution of mesoniviruses had ever been performed, we investigated whether this would extend to mesoniviruses using the RDP4 software. A full genome scan (using all detection methods implemented in RDP4) was performed, and while many minor recombination events were detected, only one potential recombination event was strongly supported by the RDP4 software, regardless of the recombination detection method used. This event seems to have been involved in the genesis of NDiV (accession number KF771866), as its genome appears to have resulted from the recombination of two distinct mesoniviruses, with NgeV (accession number MF176279) and OdoV (LC497422), or viruses very similar to them, suggested as the parental sequences. Due to the apparent mosaic nature of the NDiV sequence, it was removed from further phylogenetic analyses.

Phylogenetic reconstructions were based on the analyses of mesoniviruses genomic regions with high phylogenetic signals (Table 1). We focused our analyses on non-mosaic full-length genome sequences (dataset with the highest phylogenetic signal), as well as two others comprising different ORF-specific datasets with higher signal (ORF1b and ORF2a/S) which encode very different types of viral proteins. However, instead of analyzing the whole of the ORF1b-coding sequence, we sought to focus our analyses exclusively on the RdRp coding sequences, which not only displays high phylogenetic

signal, but especially because it is, by far, the mesonivirus genomic sequence most frequently found in public databases.

Phylogenetic reconstructions carried out using either the ORF2a/S-coding region or the full-length genomic sequence translated into similar results (Supplementary Fig. 5 for ORF2a/S and Supplementary Fig. 6 for full-length). When the current taxonomy of mesoniviruses (according to the latest update on ICTV) is superimposed to the topology of these trees and to the mesonivirus species demarcation criteria defined by Vasilakis et al. in 2014 (96.8% protein sequence identity), evident discrepancies were found when the topologies of the complete genome/S and RdRp trees were compared (compare Supplementary Fig. 5 vs. Fig. 2). Even as the trees appeared to be topologically similar, they were not identical. For example, CAVV was placed in the lineage defining the Alphamesonivirus 1 species only in the RdRp tree, and this association seemed to be supported by all PCOORD analyses. In addition, the monophyletic group that included DkNV and KPhV sequences is indicated, in the RdRp tree (and is supported by PCOORD), as sharing direct ancestry and forming a robust monophyletic clade with the lineage that clusters KSaV and BBaV, when this is not seen in the S-protein tree. These results indicate that while the mesonivirus genome or the RdRp and S regions may be used for phylogenetic analysis, some topological discrepancies are seen depending on the region used. Clearly, if tree topologies are considered to aid taxonomic decisions, even slightly different topologies may impact viral taxonomy.

Since species demarcation criteria for nidoviruses have been most focused on highly conserved RdRp aa sequences (Cowley and Walker, 2014), and since mesonivirus species demarcation have previously been focused on the analysis of concatenated regions of highly conserved domains within ORF1ab (Vasilakis et al., 2014), to evaluate how the species demarcation criteria would affect mesoniviruses classification, we focused our attention on the RdRp aa tree (Fig. 2). When phylogenetic relationships among MM were superimposed to the nomenclature scheme current defining the mesonivirus taxonomy, some clashes between tree topology and previous taxonomy assignments become apparent. KPhV and BBaV are both assigned as members of the Alphamesonivirus 1/AMV1 species and of the *Namcalivirus* subgenus in NCBI's taxonomy browser (Schoch et al., 2020), which would mean, by the obtained tree topology, that namcaliviruses are paraphyletic. However, previous studies never did place KPhV and

BBaV into that specific Alphamesonivirus species (Vasilakis et al., 2014; Wang et al., 2017; Newton et al., 2020), with Vasilakis et al. (2014) even suggesting, by PED analysis, that these two species should be considered as separate species. Indeed, our analysis did confirm this. If the minimum of 96.8% protein sequence identity defines the limit of viral species (Vasilakis et al., 2014), KPhV can never be assigned as a member of the Alphamesonivirus 1/AMV1 species. However, KPhV RdRp shares only 92% sequence identity with those of *bona fide* members of AMV1, and this suggestion is further supported by all the tree topologies obtained. Again, phylogenetic information and distance values clearly indicate KPhV and DkNV (which share 99% of RdRp sequence identity) should be members of the same (AMV3) species, confirming phylogenetic assessments in recent studies (Wang et al., 2017; Newton et al., 2020). In a similar situation, the analysis of the phylogenetic tree topologies clearly showed that BBaV should also not belong to the AMV1 species. Furthermore, when the RdRp sequences of BBaV are compared with those of AMV1 members, as mentioned above for KPhV, their RdRp share only around 90% of sequence identity. Therefore, it should not be classified as a member of AMV1. On the other hand, BBaV does seem to share common ancestry with KSaV, but both these virus' RdRp sequences form independent monophyletic clusters in phylogenetic trees (Fig. 2). While Vasilakis et al. (2014) did suggest BBaV and KSaV should be considered as separate species, Wang et al. (2017) and Newton et al. (2020) place them into the same species (AMV2). While their sequences do share high similarity, their RdRp shared only 96% identity, falling below the 96.8% cut-off value for viral species assignment. Therefore, both the RdRp phylogenetic tree topology and sequence similarity values support previous claim by Vasilakis et al. (2014) that these two viruses should be placed into two different viral species in the *Mesoniviridae* family. Since KSaV has been detected first (Vasilakis et al., 2014) and assigned to the AMV2 species, we tentatively propose that BBAV should, instead, be regarded a member of the new Alphamesonivirus 11 species (Fig. 2). Our analysis also suggested that the Odorna virus from Ghana, which remains unclassified up to the present day, should also belong to the AMV1 sublineage, as it shares over 97% of identity to other AMV1 sequences. The mesonivirus classification at the subgenus level has not yet been extensively studied and should also be reconsidered from what is currently assigned at the NCBI taxonomy browser, where the *Namcalivirus* subgenus not only contains all previously defined

mesoniviruses in the AMV1 species, but also KPhV and BBaV, while the *Karsalivirus* only contains DkNV and KSaV. We propose for a reference value of 93% of protein sequence identity (RdRp protein sequences) to be used as a reference for definition of a new subgenus in the *Mesoniviridae* family. DkNV, KPhV, BBaV and KSaV share less than 93% identity between them, and more than 93% against other MM. As such, they should all be inserted into one specific subgenus, in this case the *Karsalivirus* subgenus, while the *Namcalivirus* should only contain the species in its monophyletic clade (seen in Fig. 2, with all its sequences sharing more than 93% identity values). All the remainder sequences, Alphamesonivirus 4 (CASV), Alphamesonivirus 5 (HanaV), Alphamesonivirus 6 (OFAV), Alphamesonivirus 7 (KADV), Alphamesonivirus 8 (NseV), Alphamesonivirus 9 (MenoV) and Alphamesonivirus 10 (DKV), were represented by one single sequence each in the ML phylogenetic tree, where they appear as isolated branches, and their taxonomy classification, both at genus and subgenus level, have been reinforced by the results/findings of this study.

3.4. Analyses with other mesonivirus and virus from other families of nidoviruses

To further extend the phylogenetic analyses carried out in this work, we reconstructed the evolutionary relationships of mosquito mesonivirus, not only to other viruses in the *Nidovirales* Order, but also with the recently described mesonivirus identified in non-mosquito hosts. An initial tree was obtained using ORF1ab aa sequences (not shown), but since still no full-length ORF1ab sequence has yet been described for LtM, phylogenetic reconstruction was refined using only RdRp sequences (Supplementary Fig. 7). In the suborder *Mesnidovirineae*, mesoniviruses clearly form a stable clade that shares ancestry with the so-called Beihai Nido-like virus, the single representative of the *Medioniviridae* family, but those found in hosts other than mosquitoes (OM: AcMSV, Immeso1, LtM) are positioned between the large monophyletic clade that defines the mosquito mesonivirus lineage, as independent (not forming a cluster) sister lineages of the latter. The phylogenetic relatedness between MM and OM was not only suggested by the topology of the RdRp tree, but also by the genetic distance values obtained when OM and MM sequences were compared, indicating that OM were consistently closer to MM than to the *Medioniviridae*

Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

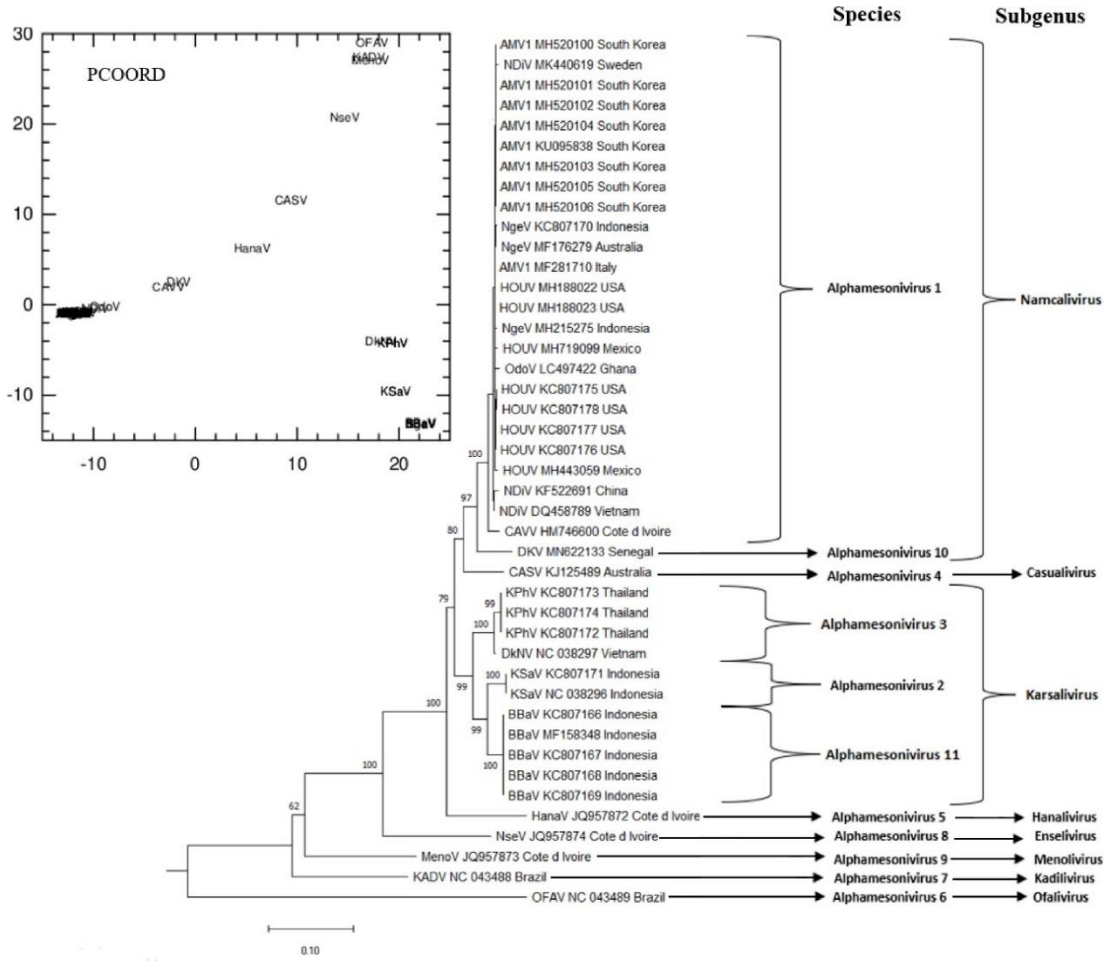


Fig. 2: Maximum likelihood tree for mosquito mesonivirus RdRp protein sequences estimated under a WAG substitution model (right panel), alongside principal coordinate analysis (left panel) carried out using PCOORD, where each sequence is identified by its corresponding abbreviation: HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV = Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus); The species and genera indicated follow the taxonomy revision proposal presented in this work.

members (only two identical sequences have been described, both from the same species, so only one is indicated in Supplementary Fig. 6). Finally, assessments of OM vs OM and OM vs MM protein sequence divergence between OM sequences, led to high divergence values, which further suggests they may correspond to the maiden members of putative new mesonivirus taxon (genera, family). These hypotheses will be investigated as further OM sequences become available.

4. Discussion

In this study, we sought to extend previously published genetic characterization data (Vasilakis et al., 2014) regarding the *Mesoniviridae* family of viruses. Over recent years, the number of mesonivirus sequences deposited in GenBank has significantly increased, which expanded the potential for new genetic analyses and phylogenetic inference analyses. While many newly described sequences clustered into predefined mesonivirus genetic lineages (like most of the sequences of the Alphamesonivirus 1 species), some were classified as totally new species (e.g., the Dianke virus). More importantly, phylogenetic reconstructions and sequence similarity calculations carried out during our study brought out new information that calls for a revision of the classification (taxonomy) of mesoniviruses.

Unlike previous reconstructions (Vasilakis et al., 2014; Chang et al., 2020) which mostly focused on the analysis of S-sequences, we performed analyses based on different sets of nucleotide and protein sequences. While the phylogenetic signals and tree topologies were calculated using multiple sequence datasets corresponding to the total coding-sequence or its ORF-specific fractions (most of which display high phylogenetic signal), special attention was devoted to phylogenetic reconstructions involving the RdRp-specific coding region, which has been extensively used for nidovirus species demarcation criteria (Cowley and Walker, 2014). Focusing on a specific dataset of sequences is important in this specific case, since our data indicate that phylogenetic reconstructions based on different genomic regions (complete-genome or ORF2A/S vs RdRp) does not result in congruent topologies, therefore confounding the establishment of clade demarcation criteria, and consequently complicates the taxonomic classification of these viruses. Therefore, we suggest that mesonivirus taxonomy should focus on the

analysis of only RdRp aa sequences, using a minimum reference value of 96.8% of protein sequence identity to define a mesonivirus species, and 93.0% of protein sequence identity as a reference value of to place mesoniviruses into a given subgenus. Not only did our analysis indicate that tree topologies and genetic diversity values at times clash with the prevailing classification scheme, the description of new viral sequences in the coming years will bring new light into the structure of the mesonivirus taxon. We proposed that Odorna (which has remained unclassified up to the present day) should become a member of the Alphamesonivirus 1 species in the *Namcalivirus* subgenus. Also, our analysis suggests that BBaV should be regarded as a candidate for new species in the *Mesoniviridae* family, tentatively named as Alphamesonivirus 11. On the other hand, the *Karsalivirus* genus, currently containing both DkNV (AMV3) and KSaV (AMV2), should also contain both BBaV (AMV11) and KPhV (AMV3). As for the remaining species, our analyses reinforced the currently accepted classification.

While taxonomic decisions based on the analysis of a small section of a viral genome (RdRp), as opposed to the use of the viral genome, may be disputed, RdRp sequences are currently the most frequently represented mesonivirus sequence in the public genomic databases. Future studies should not only focus on the *Mesoniviridae* family but also other nidoviruses and ideally should focus on obtaining full-length sequences, as our results indicate it has the highest phylogenetic signal. These studies should also combine phylogenetic, genetic distance, and statistical analyses (such as PCOORD) as complementary tools for genetic analyses.

Previous observations did indicate a worldwide distribution mesonivirus (Vasilakis et al., 2014), which we further emphasized here, with the analysis of mesonivirus sequences obtained from mosquitos collected in multiple continents. However, no signs were found regarding geographic segregation patterns. There were also no obvious signs of species-specific host restrictions, unlike what has been described for most lineages of insect-specific flaviviruses (Colmant et al., 2017). For example, like what happens to the Alphamesonivirus 1 species, most sequences have been obtained from multiple subspecies of *Culex* or *Aedes* mosquitos from multiple countries. However, the majority of the other Alphamesonivirus species are currently characterized by either one or only a few genetically close members with a similar geographic origin, which confounds the identification of possible geographic or host-range limits.

The detailed analysis of mesonivirus genomic features confirmed that newly described MM sequences conform, in general terms, to the genetic organization previously defined for mesoniviruses. When compared to other nidovirus, mesonivirus have shorter genomes (the only exception being the *Arteriviridae* family, with even shorter genomes) as well as a lower number of identifiable ORFs. The analysis of recently described mesonivirus sequences indicated the presence of a sequence insertion block in ORF1a similar to that reported previously (Vasilakis et al., 2014) in DkNV. Nidovirus genome expansions have been previously reported (Lauber et al., 2013), but the specific functional role of the ORF1a insertion blocks remains unclear. Our analysis suggested it has coding capacity, though its putative product is of unknown function. Other than function of the products of the readily identifiable ORF1a and ORF1b, the putative functions of the other ORFs found in the MM viral sequences analyzed remains elusive. However, remote homology detection and some of their biochemical features suggest two of them encode a nucleocapsid and a viral envelope glycoprotein. Altogether, mesoniviruses are characterized by low amino acid sequence diversity (by entropy assessment), as well as a lower number of non-synonymous substitutions (by calculation of ω values), especially when compared to other nidoviruses.

We executed the first phylogenetic reconstruction with multiple meso-like virus isolated from non-mosquito hosts to elucidate how they fare in phylogenetic relationships into the whole *Nidovirales* Order radiation. They were all classified as members of the *Nidovirales* Order based on its genomic structure, amino acid sequence identity and phylogenetic analysis, expanding our knowledge on the host range of mesonivirus, previously only reported in mosquitos. Although they could tentatively be classified as mesonivirus, their sequence identity with mosquito mesonivirus and virus from other close families, like the *Medioniviridae* and *Coronaviridae* family, is quite low. Even between themselves there is high sequence discrepancy, and there are significant differences in their genomic structure. While the more conserved regions (both ORF1a and ORF1b) look to be nearly identical to other mesonivirus (including its putative proteome characteristics), the remaining ORFs, which should correspond to structural protein coding regions, did not found any similarity searches, with no known function or domain found as well. Further studies are needed, as more non-mosquito meso-like virus are identified in the future, to evaluate whether these new viruses could indeed be

clustered with viruses of the *Mesoniviridae* family, or even shape a new family. Also, this does corroborate past studies that hinted at the evolution of nidovirus in arthropods and consequent spread into other group of hosts, including vertebrates (Nga et al., 2011), which may happen again with mesonivirus. Coronavirus from completely different hosts (like bats and equines) share low sequence identity, which also happens between mosquito mesonivirus and other mesonivirus (with non-mosquito hosts).

In contrast with the high phylogenetic signal values associated with of most datasets of mesonivirus sequences, assessment of sequence divergence through time using root-to-tip analysis, systematically indicated, for all datasets used, that there is still insufficient data in the public databases to possibly support a phylogeographic reconstruction of the evolution of mesoniviruses. This result is most probably the consequence of a poor range of sampling time for existing sequences, which would negatively impact the sequences' temporal signal. In fact, even though some sequences (from BBaV, KPhV and NgeV) were detected in mosquito collections from the early 1980's, the remainder have been mostly obtained from mosquitos collected in the last 10 to 15 years. Therefore, phylogeographic reconstructions that would disclose the geographic distribution of mesoniviruses over time still awaits that more diverse assemblages of heterochronous mesonivirus sequences become available in the near future. Mesoniviruses look to be an ever-expanding and unique group of viruses in the *Nidovirales* Order, with more information being obtained as new sequences are identified. Even their stance as insect-specific viruses could no longer hold true, as more hosts continue to be recognized (if they indeed end up being classified as virus in the *Mesoniviridae* family). Studies like these should continue to be executed in the future. Their potential to be developed as biological control agents, which have been identified in similar viruses (Goenaga et al., 2020), also remains unclear and is an important area for future investigation.

Supporting information

Supplementary Fig. 1: Schematic representation of nucleotide sequences for each group of mesonivirus and for other virus from different families in the *Nidovirales* Order, with different ORFs identified; * - Nucleotide sequence for the Odorna virus seems incomplete and contains no further information apart from the one present here; RFS: ribosomal frameshift elements; HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV =

Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus; SARS-CoV-2 = Severe acute respiratory syndrome coronavirus 2 (*Coronaviridae*; MT997203); SheV = Simian hemorrhagic encephalitis virus (*Arteriviridae*; NC_038293); FmN = Fathead minnow nidovirus (*Tobaniviridae;* NC_038295); BlN = Botrylloides leachii nidovirus (*Medionivirineae*; MK956105).

Supplementary Fig. 2: Heat map representing intersequence genetic diversity of mesonivirus. Representative tree (maximum likelihood, WAG model) based on RdRp aa sequences (sequences identifiable in Supplementary Table 1), and Z-Scores were obtained based on pairwise evolutionary distances obtained on MegaX.

Supplementary Fig. 3: Intragroup genetic diversity of mesonivirus. Representative RdRp tree (maximum likelihood, WAG model) based on the analysis of alignments of RdRp primary sequences. For each species, sequence identification follows the nomenclature indicated in Supplementary Table 1, followed by number of sequences for each clade; box-and-whisker graphs are used to plot distributions of pairwise evolutionary distances of three different sets: between mesoniviruses from the same species (Alphamesonivirus 1/AMV1), between all mosquito mesoniviruses (MM), and between all mosquito mesoniviruses (MM) and other mesoniviruses (OM).

Supplementary Fig. 4: Entropy calculations based the Shannon function (Shannon entropy-one) applied on alignments of ORF1a protein sequences from different families in the *Nidovirales* Order.

Supplementary Fig. 5: Principal coordinate analysis carried (left panel) out for mosquito mesonivirus S protein coding sequences. Each sequence is identified by the sequence abbreviation they belong to (HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV = Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus). A maximum likelihood tree (right panel), estimated under a

WAG substitution model, is also shown, while displaying the taxonomy revision proposal presented in this work.

Supplementary Fig. 6: Maximum likelihood tree for mosquito mesonivirus full-length sequences. Each sequence is identified by the sequence abbreviation they belong to (HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV = Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus).

Supplementary Fig. 7: Maximum likelihood tree for protein sequences (RdRp coding region) of virus from different families in the *Nidovirales* Order.

CRediT authorship contribution statement

Paulo Morais: Data curation, Formal analysis, Visualization. **Nídia S. Trovão:** Writing – review & editing. **Ana B. Abecasis:** Writing – review & editing. **Ricardo Parreira:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work received financial support from the Global Health and Tropical Medicine Center, which is funded through FCT contract UID/ Multi/04413/2013. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2022.198727.

References

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S., 2016. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 44 (W1), W147–W153. https://doi.org/10.1093/nar/gkw419.

Blitvich, B.J., Firth, A.E., 2015. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. Viruses 7 (4), 1927–1959. https://doi.org/10.3390/v7041927.

Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.Y., Perry, B.W., Castoe, T.A., Robertson, D.L., 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. 5 (11), 1408–1417. https://doi.org/10.1038/s41564-020-0771-4.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17 (4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chang, T., Guo, M., Zhang, W., Niu, J., Wang, J.J., 2020. First report of a mesonivirus and its derived small RNAs in an aphid species aphis citricidus (Hemiptera: Aphididae), implying viral infection activity. J. Insect Sci. 20 (2), 1–4. https://doi.org/10.1093/jisesa/ieaa022.

Charles, J., Tangudu, C.S., Hurt, S.L., Tumescheit, C., Firth, A.E., Garcia-Rejon, J.E., Blitvich, B.J., 2018. Detection of novel and recognized RNA viruses in mosquitoes from the Yucatan Peninsula of Mexico using metagenomics and characterization of their *in vitro* host ranges. J. Gen. Virol. 99 (12), 1729–1738. https://doi.org/10.1099/jgv.0.001165.

Chiapello, M., Bosco, L., Ciuffo, M., Ottati, S., Salem, N., Rosa, C., Turina, M., 2021. Complexity and local specificity of the virome associated with tospovirus-transmitting thrips species. J. Virol. 95 (21) https://doi.org/10.1128/jvi.00597-21.

Colmant, A.M.G., Hobson-Peters, J., Bielefeldt-Ohmann, H., van den Hurk, A.F., Hall- Mendelin, S., Chow, W.K., Hall, R.A., 2017. A new clade of insect-specific

flaviviruses from Australian anopheles mosquitoes displays species-specific host restriction. MSphere 2 (4), 155–166. https://doi.org/10.1128/mSphere.00262-17.

Cowley, J.A., Walker, P.J., 2014. Molecular biology and pathogenesis of roniviruses. Nidoviruses 361–377. https://doi.org/10.1128/9781555815790.ch24.

Diagne, M.M., Gaye, A., Ndione, M.H.D., Faye, M., Fall, G., Dieng, I., Sall, A.A., 2020. Dianke virus: a new mesonivirus species isolated from mosquitoes in eastern Senegal. Virus Res. 275 (June 2019), 197802 https://doi.org/10.1016/j.virusres.2019.197802.

Goenaga, S., Goenaga, J., Boaglio, E.R., Enria, D.A., Levis, S.D.C., 2020. Superinfection exclusion studies using west nile virus and culex flavivirus strains from Argentina. Mem. Inst. Oswaldo Cruz. 115 (5), 1–5. https://doi.org/10.1590/0074-02760200012.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Research 117, 17–37. https://doi.org/10.1016/j.virusres.2006.01.017.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95-98-NT. Nucleic Acids Symp. Ser. 41, 95–98.

Hon, C.C., Lam, T.Y., Shi, Z.L., Drummond, A.J., Yip, C.W., Zeng, F., Leung, F.C.C, 2008. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. J. Virol. 82 (4), 1819–1826. https://doi.org/10.1128/JVI.01926-07.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30 (14), 772–780. https://doi.org/10.1093/molbev/mst010.

Kosakovsky Pond, S.L., Frost, S.D.W, 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21 (10), 2531–2533. https://doi.org/10.1093/bioinformatics/bti320.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35 (6), 1547–1549. https://doi.org/10.1093/molbev/msy096.

Lai, C., Shih, T., Ko, W., Tang, H., Hsueh, P., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. Int. J. Antimicrob. Agents 55 (3), 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924.

Lauber, C., Goeman, J.J., de Parquet, M.C., Thi Nga, P., Snijder, E.J., Morita, K., Gorbalenya, A.E., 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS Pathog. 9 (7) https://doi.org/10.1371/journal.ppat.1003500.

Letunic, I., Bork, P., 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46 (D1), D493–D496. https://doi.org/10.1093/nar/gkx922.

Marklewitz, M., Zirkel, F., Rwego, I.B., Heidemann, H., Trippner, P., Kurth, A., Junglen, S., 2013. Discovery of a unique novel clade of mosquito-associated bunyaviruses. J. Virol. 87 (23), 12850–12865. https://doi.org/10.1128/jvi.01862- 13.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1 (1), 1–5. https://doi.org/10.1093/ve/vev003.

Nei, M., Gojoborit, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426. https://doi.org/10.1093/oxfordjournals.molbev.a040410.

Newton, N.D., Colmant, A.M.G., O'Brien, C.A., Ledger, E., Paramitha, D., Bielefeldt- Ohmann, H., Hobson-Peters, J., 2020. Genetic, morphological and antigenic relationships between mesonivirus isolates from Australian mosquitoes and evidence for their horizontal transmission. Viruses 12 (10). https://doi.org/10.3390/v12101159.

Nga, P.T., de Parquet, M.C., Lauber, C., Parida, M., Nabeshima, T., Yu, F., Gorbalenya, A. E., 2011. Discovery of the first insect nidovirus, a missing evolutionary

link in the emergence of the largest RNA virus genomes. PLoS Pathog. 7 (9) https://doi.org/10.1371/journal.ppat.1002215.

Rambaut, A., Lam, T.T., Carvalho, L.M., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2 (1) https://doi.org/10.1093/ve/vew007.

Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Karsch-Mizrachi I. et al. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database, 2020. https://doi.org/10.1093/database/baaa062.

Schmidt, H.A., Strimmer, K., Vingron, M., Von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18 (3), 502–504. https://doi.org/10.1093/bioinformatics/18.3.502.

Strimmer, K., Von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Nat. Acad. Sci. USA 94 (13), 6815–6819. https://doi.org/10.1073/pnas.94.13.6815.

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44 (W1), W232–W235. https://doi.org/10.1093/nar/gkw256.

Vasilakis, N., Guzman, H., Firth, C., Forrester, N.L., Widen, S.G., Wood, T.G., Tesh, R.B., 2014. Mesoniviruses are mosquito-specific viruses with extensive geographic distribution and host range. Virol. J. 11 (1), 1–12. https://doi.org/10.1186/1743-422X-11-97.

Wang, Y., Xia, H., Zhang, B., Liu, X., Yuan, Z., 2017. Isolation and characterization of a novel mesonivirus from Culex mosquitoes in China. Virus Res. 240 (June), 130–139. https://doi.org/10.1016/j.virusres.2017.08.001.

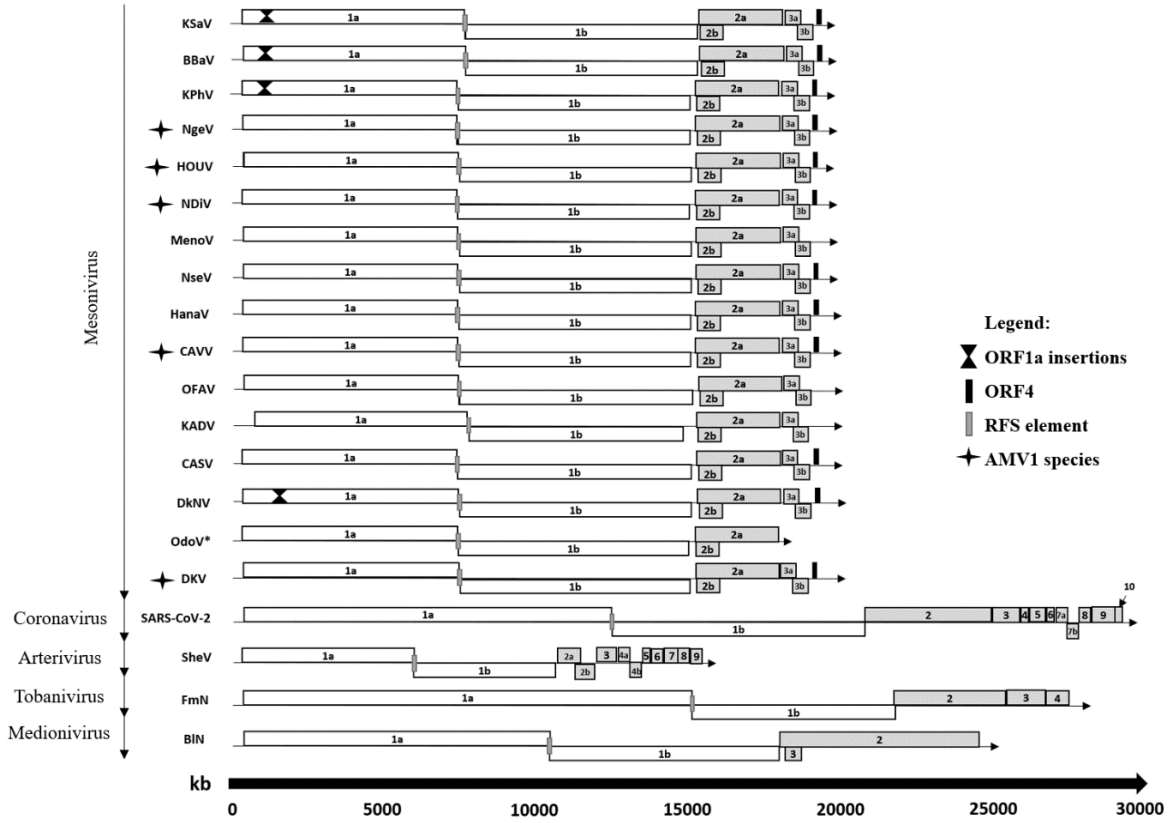
Warrilow, D., Watterson, D., Hall, R.A., Davis, S.S., Weir, R., Kurucz, N., Hobson- Peters, J., 2014. A new species of mesonivirus from the northern territory, Australia. PLoS ONE 9 (3). https://doi.org/10.1371/journal.pone.0091103.

Zimmermann, L., Stephens, A., Nam, S.Z., Rau, D., Kübler, J., Lozajic, M., Alva, V., 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J. Mol. Biol. 430 (15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.

Zirkel, F., Kurth, A., Quan, P.L., Briese, T., Ellerbrok, H., Pauli, G., Junglen, S., 2011. An insect nidovirus emerging from a primary tropical rainforest. MBio 2 (3), 21–23. https://doi.org/10.1128/mBio.00077-11.

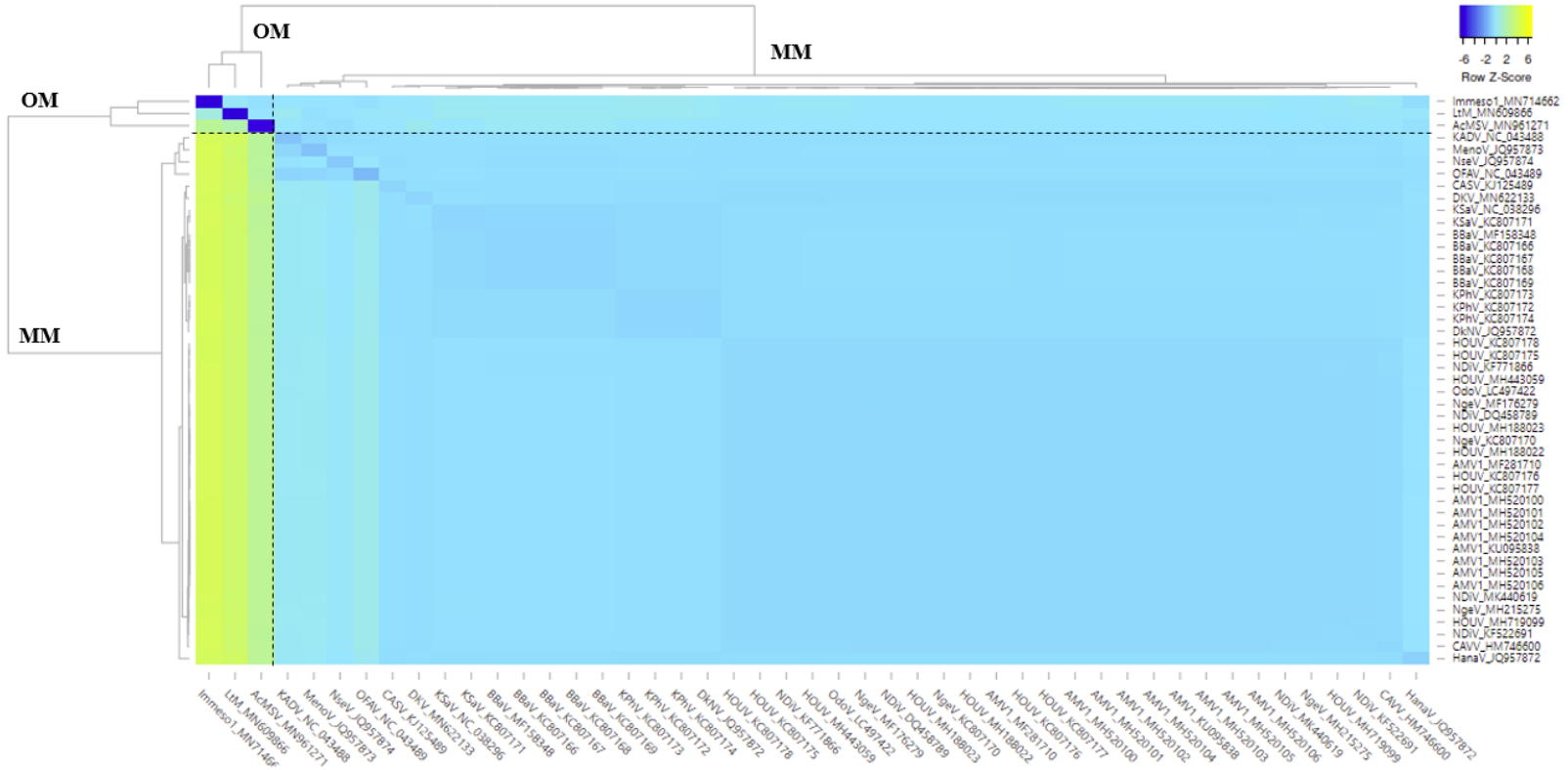
Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Supplementary Fig. 1: Schematic representation of nucleotide sequences for each group of mesonivirus and for other virus from different families in the *Nidovirales* Order, with different ORFs identified; * - Nucleotide sequence for the Odorna virus seems incomplete and contains no further information apart from the one present here; RFS: ribosomal frameshift elements; HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV = Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus; SARS-CoV-2 = Severe acute respiratory syndrome coronavirus 2 (*Coronaviridae*; MT997203); SheV = Simian hemorrhagic encephalitis virus (*Arteriviridae*; NC_038293); FmN = Fathead minnow nidovirus (*Tobaniviridae;* NC_038295); BlN = Botrylloides leachii nidovirus (*Medionivirineae*; MK956105).



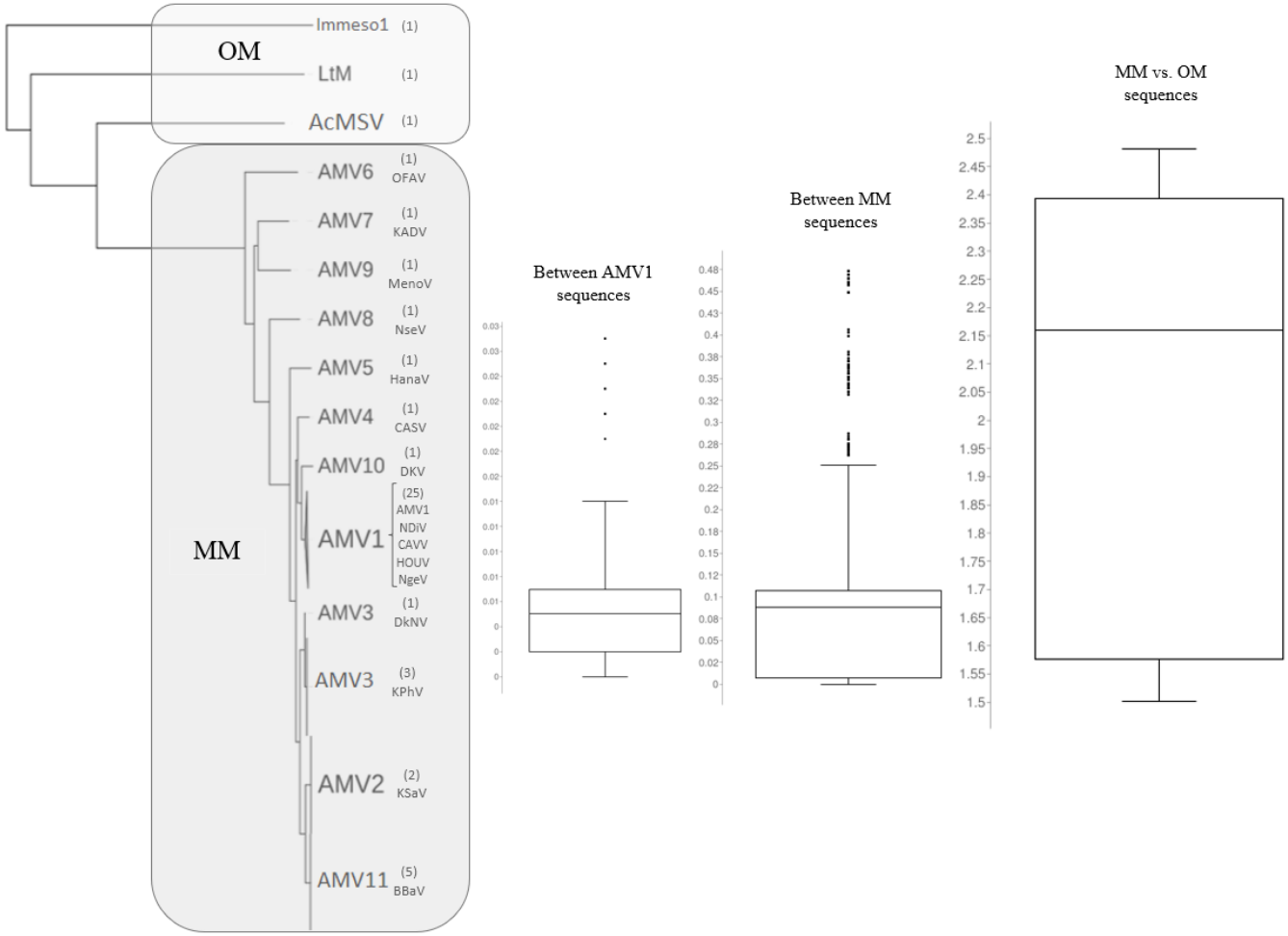
Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Supplementary Fig. 2: Heat map representing intersequence genetic diversity of mesonivirus. Representative tree (maximum likelihood, WAG model) based on RdRp aa sequences (sequences identifiable in Supplementary Table 1), and Z-Scores were obtained based on pairwise evolutionary distances obtained on MegaX.

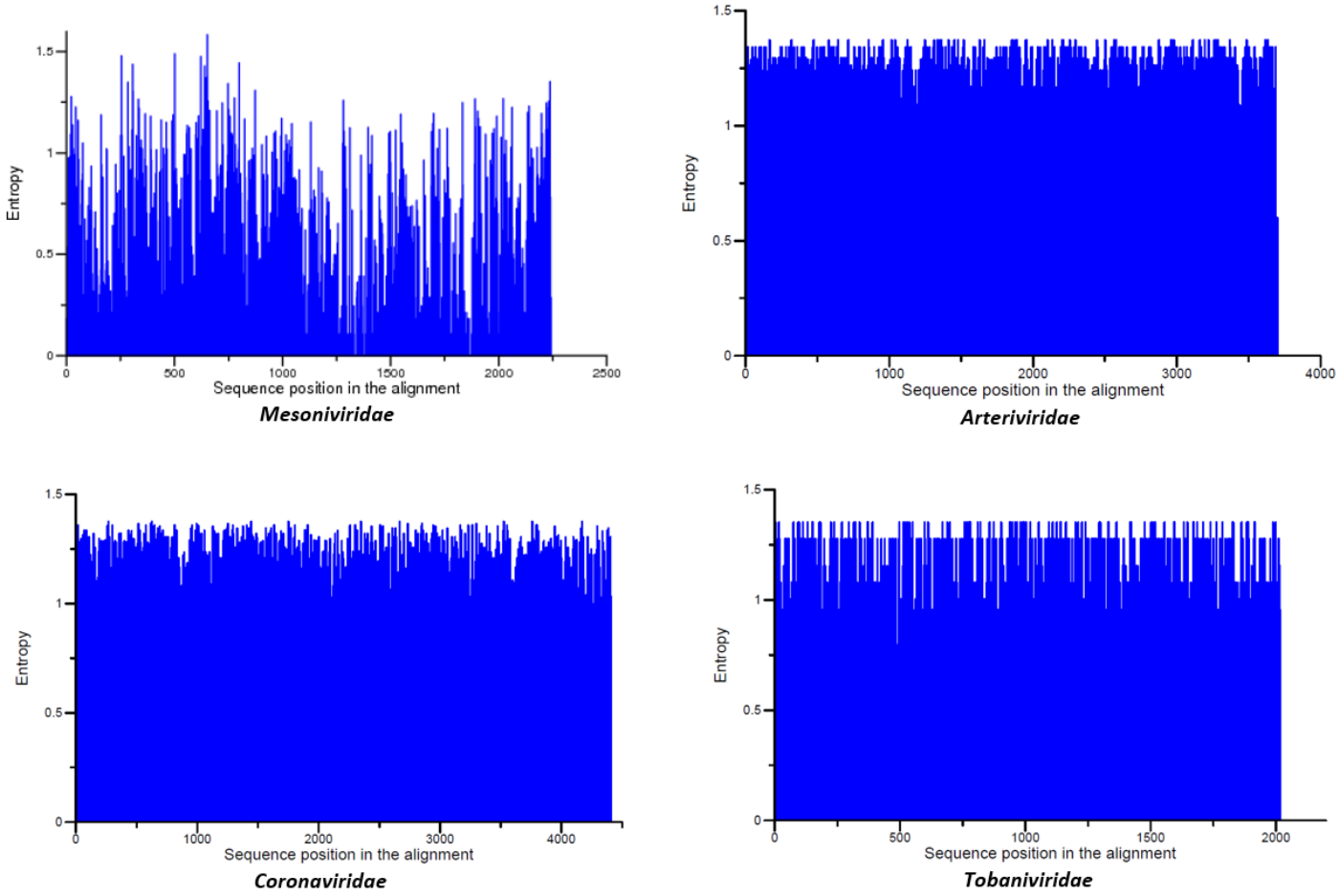


Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

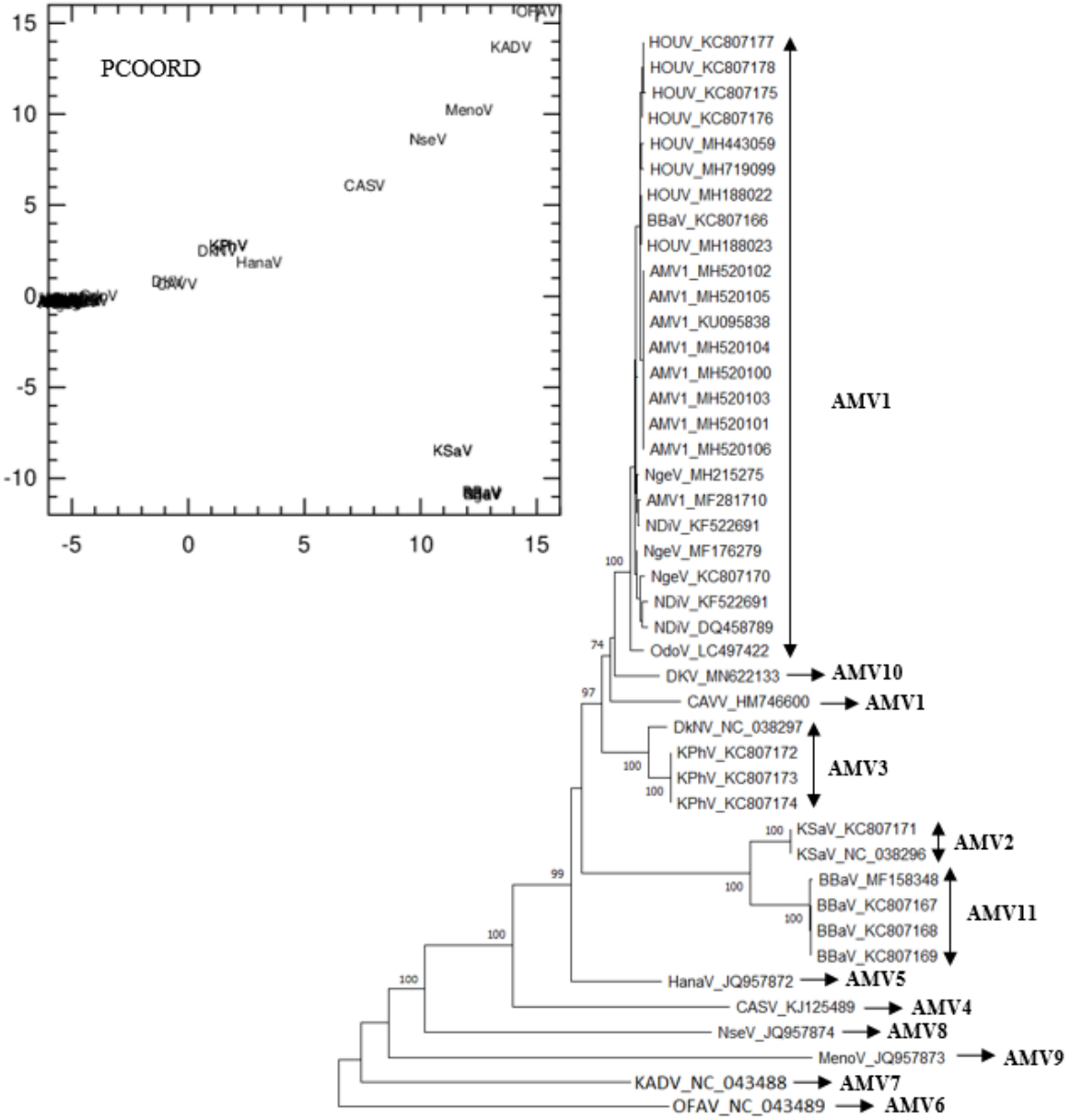
Supplementary Fig. 3: Intragroup genetic diversity of mesonivirus. Representative RdRp tree (maximum likelihood, WAG model) based on the analysis of alignments of RdRp primary sequences. For each species, sequence identification follows the nomenclature indicated in Supplementary Table 1, followed by number of sequences for each clade; box-and-whisker graphs are used to plot distributions of pairwise evolutionary distances of three different sets: between mesoniviruses from the same species (Alphamesonivirus 1/AMV1), between all mosquito mesoniviruses (MM), and between all mosquito mesoniviruses (MM) and other mesoniviruses (OM).



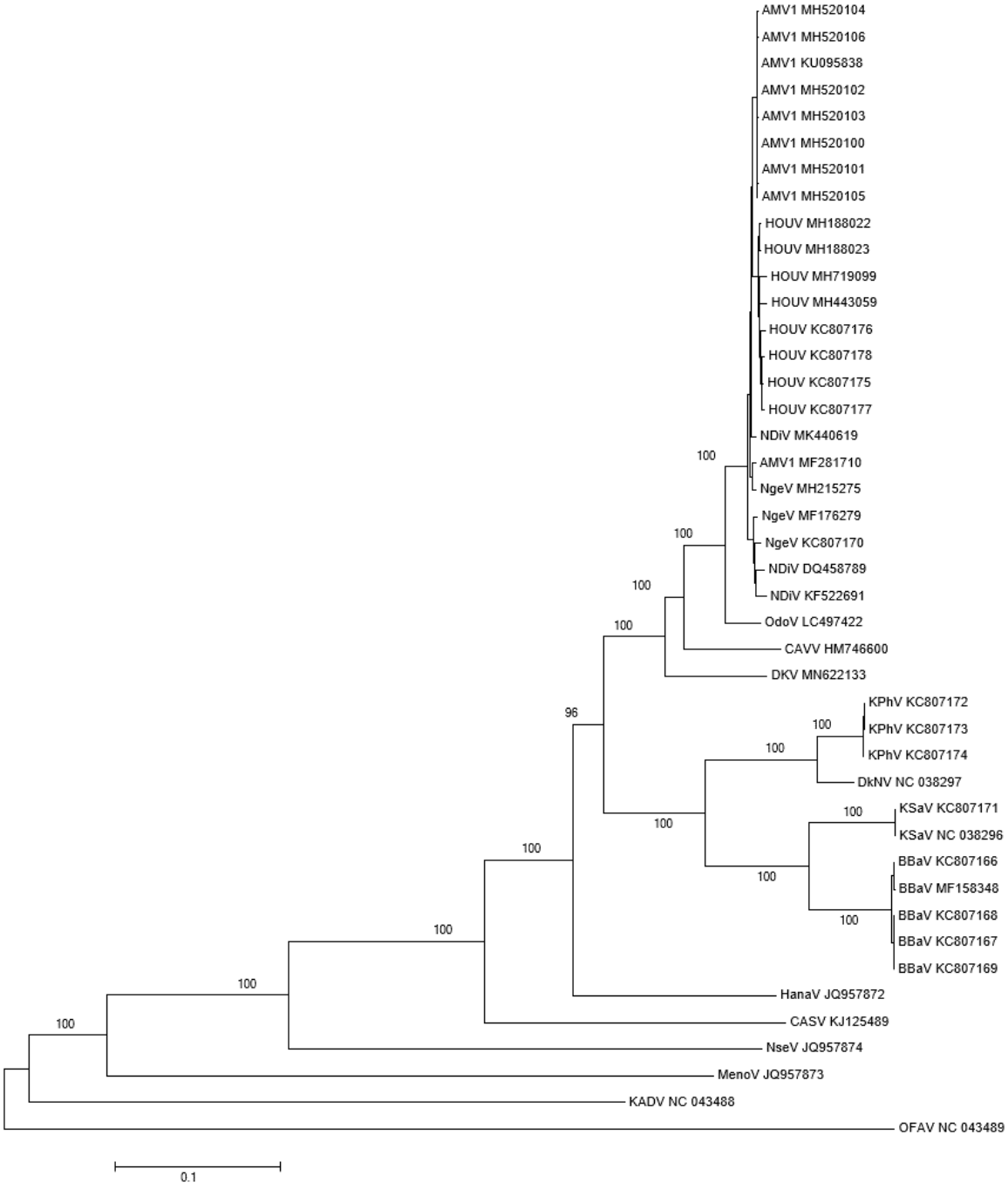
Supplementary Fig. 4: Entropy calculations based the Shannon function (Shannon entropy-one) applied on alignments of ORF1a protein sequences from different families in the *Nidovirales* Order.



Supplementary Fig. 5: Principal coordinate analysis carried (left panel) out for mosquito mesonivirus S protein coding sequences. Each sequence is identified by the sequence abbreviation they belong to (HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV = Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus). A maximum likelihood tree (right panel), estimated under a WAG substitution model, is also shown, while displaying the taxonomy revision proposal presented in this work.

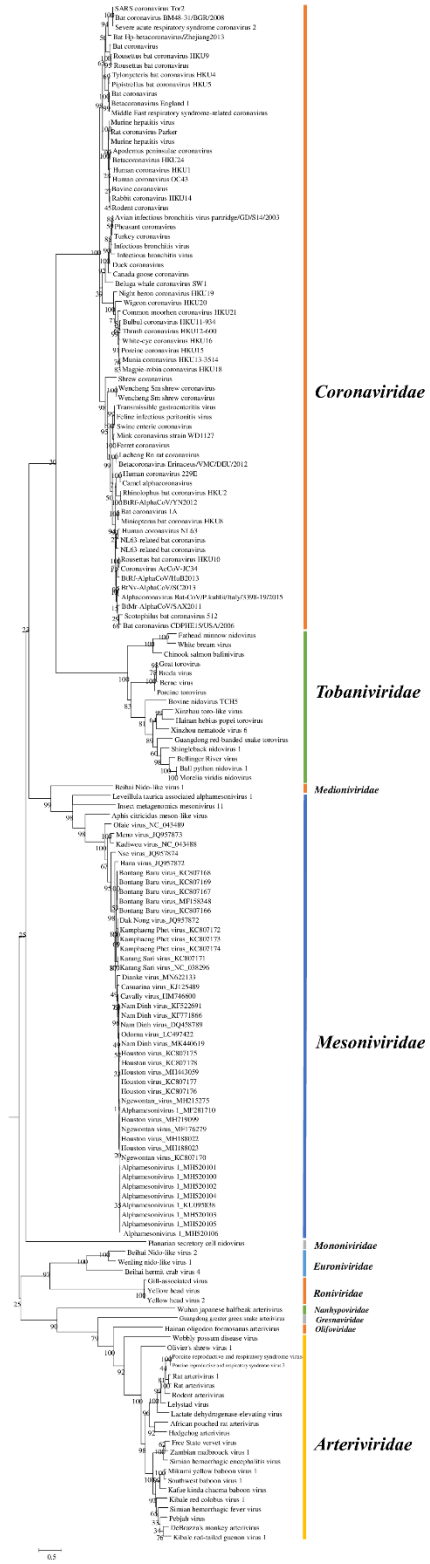


Supplementary Fig. 6: Maximum likelihood tree for mosquito mesonivirus full-length sequences. Each sequence is identified by the sequence abbreviation they belong to (HOUV = Houston virus; AMV1 = Alphamesonivirus 1; NDiV = Nam Dinh virus; NgeV = Ngewontan virus; OdoV = Odorna virus; CAVV = Cavally virus; DKV = Dianke virus; HanaV = Hana virus; BBaV = Bontang Baru virus; KSaV = Karang Sari virus; KPhV = Kamphaeng Phet virus; DkNV = Dak Nong virus; CASV = Casuarina virus; NseV = Nse virus; KADV = Kadiweu virus; MenoV = Meno virus; OFAV = Ofaie virus).



Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Supplementary Fig. 7: Maximum likelihood tree for protein sequences (RdRp coding region) of virus from different families in the *Nidovirales* Order.



Supplementary Table 1: Mesonivirus sequences (*Alphamesonivirus* genus) used in the present study.

Accession number	Subgenus	Species	Nomenclature	Abbreviation	Host species	Geographical origin	Collection Year
MH520101	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2016
MH520100	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2016
MH520102	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2016
MH520103	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2016
MH520104	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2016
MH520105	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex tritaeniorhynchus*	South Korea	2016
MH520106	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2016
MF281710	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex sp.*	Italy	2008
KU095838	*Namcalivirus*	AMV1	Alphamesonivirus 1	AMV1	*Culex pipiens*	South Korea	2012
KC807166	*Karsalivirus*	AMV11	Bontang Baru virus	BBaV	*Culex vishnui*	Indonesia	1981
KC807167	*Karsalivirus*	AMV11	Bontang Baru virus	BBaV	*Culex tritaeniorhynchus*	Indonesia	1981
KC807168	*Karsalivirus*	AMV11	Bontang Baru virus	BBaV	*Culex vishnui*	Indonesia	1981
KC807169	*Karsalivirus*	AMV11	Bontang Baru virus	BBaV	*Culex vishnui*	Indonesia	1981
MF158348	*Karsalivirus*	AMV11	Bontang Baru virus	BBaV	*Culex vishnui*	Indonesia	1981
KJ125489	*Casualivirus*	AMV4	Casuarina virus	CASV	*Coquillettidia xanthogaster*	Australia	2010
HM746600	*Namcalivirus*	AMV1	Cavally virus	CAVV	*Aedes harrisoni*	Cote d'Ivoire	2004
NC_038297	*Karsalivirus*	AMV3	Dak Nong virus	DkNV	*Culex tritaeniorhynchus*	Vietnam	2007
MN622133	*Namcalivirus*	AMV10	Dianke virus	DKV	*Aedes albopictus*	Senegal	2012
JQ957872	*Hanalivirus*	AMV5	Hana virus	HanaV	*Culex sp.*	Cote d'Ivoire	2004
MH719099	*Namcalivirus*	AMV1	Houston virus	HOUV	*Culex quinquefasciatus*	Mexico	2008

KC807175	*Namcalivirus*	AMV1	Houston virus	HOUV	*Culex quinquefasciatus*	USA	2004
KC807176	*Namcalivirus*	AMV1	Houston virus	HOUV	*Culex quinquefasciatus*	USA	2004
KC807177	*Namcalivirus*	AMV1	Houston virus	HOUV	*Aedes albopictus*	USA	2010
KC807178	*Namcalivirus*	AMV1	Houston virus	HOUV	*Aedes albopictus*	USA	2010
MH188023	*Namcalivirus*	AMV1	Houston virus	HOUV	*Culex sp.*	USA	2016
MH188022	*Namcalivirus*	AMV1	Houston virus	HOUV	*Culex sp.*	USA	2016
MH443059	*Namcalivirus*	AMV1	Houston virus	HOUV	*Culex quinquefasciatus*	Mexico	2017
NC_043488	*Kadilivirus*	AMV7	Kadiweu virus	KADV	*Culex sp.*	Brazil	2010
KC807172	*Karsalivirus*	AMV3	Kamphaeng Phet virus	KPhV	Unknown mosquito	Thailand	1984
KC807173	*Karsalivirus*	AMV3	Kamphaeng Phet virus	KPhV	Unknown mosquito	Thailand	1984
KC807174	*Karsalivirus*	AMV3	Kamphaeng Phet virus	KPhV	Unknown mosquito	Thailand	1984
KC807171	*Karsalivirus*	AMV2	Karang Sari virus	KSaV	*Culex vishnui*	Indonesia	1981
NC_038296	*Karsalivirus*	AMV2	Karang Sari virus	KSaV	*Culex vishnui*	Indonesia	1981
JQ957873	*Menolivirus*	AMV9	Meno virus	MenoV	*Uranotaenia chorleyi*	Cote d'Ivoire	2004
MK440619	*Namcalivirus*	AMV1	Nam Dinh virus	NDiV	*Culex sp.*	Sweden	2009
DQ458789	*Namcalivirus*	AMV1	Nam Dinh virus	NDiV	*Culex sp.*	Vietnam	2002
KF522691	*Namcalivirus*	AMV1	Nam Dinh virus	NDiV	*Culex pipiens quinquefasciatus*	China	2011
KF771866	*Namcalivirus*	AMV1	Nam Dinh virus	NDiV	Unknown mosquito	China	2009
KC807170	*Namcalivirus*	AMV1	Ngewontan virus	NgeV	*Culex vishnui*	Indonesia	1981
MF176279	*Namcalivirus*	AMV1	Ngewontan virus	NgeV	*Culex australicus*	Australia	2015
KC807170	*Namcalivirus*	AMV1	Ngewontan virus	NgeV	*Culex vishnui*	Indonesia	1981
JQ957874	*Enselivirus*	AMV8	Nse virus	NseV	*Culex nebulosus*	Cote d'Ivoire	2004
LC497422	*Namcalivirus*	AMV1	Odorna virus	OdoV	*Aedes aegypti*	Ghana	2016

NC_043489	*Ofalivirus*	AMV6	Ofaie virus	OFAV	*Mansonia sp.*	Brazil	2010
MN961271	unclassified	-	Aphis citricidus meson-like virus	AcMSV	*Aphis citricidus*	China	2017
MN714662	unclassified	-	Insect metagenomics mesonivirus 1	Immeso1	*Thrips tabaci*	Italy	2018
MN609866	unclassified	-	Leveillula taurica associated alphamesonivirus 1	LtM	*Leveillula taurica*	Italy	2018

Supplementary Table 2: Complete set of virus sequences from the *Nidovirales* Order used in the present study.

Accession number	Family	Genus	Species	Host species	Geographical origin	Collection Year
NC_026439	*Arteriviridae*	*Lambdaarterivirus*	Forest pouched giant rat arterivirus	*Cricetomys emini*	Cameroon	2010
NC_035127	*Arteriviridae*	*Muarterivirus*	Olivier's shrew virus 1	*Crocidura olivieri guineensis*	Guinea	2016
NC_002532.2	*Arteriviridae*	*Alphaarterivirus*	Equine arteritis virus	Mare	USA	1953
MH155201	*Arteriviridae*	*Deltaarterivirus*	Simian hemorrhagic fever virus	Simian	USA	1964
NC_001639.1	*Arteriviridae*	*Gammaarterivirus*	Lactate dehydrogenase-elevating virus	Mouse	USA	1960
NC_038293	*Arteriviridae*	*Epsilonarterivirus*	Simian hemorrhagic encephalitis virus	*Macaca mulatta*	Russia	1964
NC_048209	*Arteriviridae*	*Epsilonarterivirus*	Zambian malbrouck virus	*Chlorocebus sp.*	Zambia	2015
NC_029992	*Arteriviridae*	*Epsilonarterivirus*	Free State vervet virus	*Chlorocebus sp.*	South Africa	2010
NC_034455	*Arteriviridae*	*Zetaarterivirus*	Kibale red colobus virus 2	*Piliocolobus tephrosceles*	Uganda	2012
NC_033553	*Arteriviridae*	*Zetaarterivirus*	Kibale red colobus virus 1	*Piliocolobus tephrosceles*	Uganda	2012
NC_026509	*Arteriviridae*	*Iotaarterivirus*	DeBrazza monkey arterivirus	*Cercopithecus neglectus*	Cameroon	2010
NC_038292	*Arteriviridae*	*Iotaarterivirus*	Kibale red-tailed guenon virus 1	Red-tailed guenon	Uganda	2010

NC_027124	*Arteriviridae*	*Iotaarterivirus*	Pebjah virus	Rhesus macaque	USA	1989
KT895940	*Arteriviridae*	*Thetaarterivirus*	Kafue kinda chacma baboon virus	*Papio kindae x ursinus*	Zambia	2014
NC_025112	*Arteriviridae*	*Thetaarterivirus*	Mikumi yellow baboon virus 1	*Papio cynocephalus*	Tanzania	1986
MN642099	*Arteriviridae*	*Betaarterivirus*	Porcine reproductive and respiratory syndrome virus 2 (PRRSV-2)	*Sus scrofa*	China	2015
NC_043487.1	*Arteriviridae*	*Betaarterivirus*	Lelystad virus	Pig	Unknown	Unknown
NC_028963	*Arteriviridae*	*Betaarterivirus*	Rat arterivirus 1	*Myodes rufocanus*	China	2014
NC_026811	*Arteriviridae*	*Kappaarterivirus*	Wobbly possum disease virus	*Trichosurus vulpecula*	New Zealand	1995
NC_048210	*Arteriviridae*	unclassified	Rodent arterivirus	*Eothenomys inez*	China	2014
NC_025113	*Arteriviridae*	unclassified	Southwest baboon virus 1	*Papio anubis*	USA	2013
NC_001961	*Arteriviridae*	unclassified	Porcine reproductive and respiratory syndrome virus (PRRSV)	*Sus scrofa*	Unknown	Unknown
MT415062	*Arteriviridae*	unclassified	Hedgehog arterivirus	*Erinaceus europaeus*	United Kingdom	2019
NC_046958	*Olifoviridae*	*Kukrinivirus*	Hainan oligodon formosanus arterivirus	*Oligodon formosanus*	China	2018
NC_022103	*Coronaviridae*	*Alphacoronavirus*	Bat coronavirus CDPHE15	*Myotis lucifugus*	USA	2006
NC_028814	*Coronaviridae*	*Alphacoronavirus*	BtRf-AlphaCoV/HuB2013	*Rhinolophus ferrumequinum*	China	2013
MT438696	*Coronaviridae*	*Alphacoronavirus*	Human coronavirus 229E	*Homo sapiens*	USA	2016
NC_034972	*Coronaviridae*	*Alphacoronavirus*	Coronavirus AcCoV-JC34	*Apodemus chevrieri*	China	2011
NC_032730	*Coronaviridae*	*Alphacoronavirus*	Lucheng Rn rat coronavirus	*Rattus norvegicus*	China	2013
NC_028752	*Coronaviridae*	*Alphacoronavirus*	Camel alphacoronavirus	Camel	Saudi Arabia	2015
EU420138	*Coronaviridae*	*Alphacoronavirus*	Miniopterus bat coronavirus 1	*Miniopterus*	Hong Kong	2004
NC_028833	*Coronaviridae*	*Alphacoronavirus*	BtNv-AlphaCoV/SC2013	*Nyctalus velutinus*	China	2013

NC_010437	*Coronaviridae*	*Alphacoronavirus*	Bat coronavirus 1A	Bat	Hong Kong	Unknown
NC_028824	*Coronaviridae*	*Alphacoronavirus*	BtRf-AlphaCoV/YN2012	*Rhinolophus ferrumequinum*	China	2012
NC_048216	*Coronaviridae*	*Alphacoronavirus*	NL63-related bat coronavirus	*Triaenops afer*	Kenya	2010
NC_002306	*Coronaviridae*	*Alphacoronavirus*	Feline infectious peritonitis virus	Feline	USA	Unknown
NC_009657	*Coronaviridae*	*Alphacoronavirus*	Scotophilus bat coronavirus 512	Bat	Unknown	2005
NC_009988	*Coronaviridae*	*Alphacoronavirus*	Bat coronavirus HKU2	Bat	Unknown	Unknown
NC_032107	*Coronaviridae*	*Alphacoronavirus*	NL63-related bat coronavirus	*Triaenops afer*	Kenya	2010
NC_030292	*Coronaviridae*	*Alphacoronavirus*	Ferret coronavirus	*Mustela putorius*	Netherlands	2010
NC_028806	*Coronaviridae*	*Alphacoronavirus*	Swine enteric coronavirus	Pig	Italy	2009
NC_028811	*Coronaviridae*	*Alphacoronavirus*	BtMr-AlphaCoV/SAX2011	*Myotis ricketti*	China	2011
NC_035191	*Coronaviridae*	*Alphacoronavirus*	Wencheng Sm shrew coronavirus	*Suncus murinus*	China	2015
NC_009657	*Coronaviridae*	*Alphacoronavirus*	Scotophilus bat coronavirus 512	*Scotophilus kuhlii*	China	2005
NC_046964	*Coronaviridae*	*Alphacoronavirus*	Alphacoronavirus Bat-CoV/P.kuhlii/Italy/3398-19/2015	*Pipistrellus kuhlii*	Italy	2015
NC_038861	*Coronaviridae*	*Alphacoronavirus*	Transmissible gastroenteritis virus	Pig	Unknown	Unknown
NC_005831	*Coronaviridae*	*Alphacoronavirus*	Human coronavirus NL63	*Homo sapiens*	Netherlands	2003
FJ938051	*Coronaviridae*	*Alphacoronavirus*	Feline coronavirus RM	Cat	USA	2002
NC_023760	*Coronaviridae*	*Alphacoronavirus*	Mink coronavirus	*Mustela vison*	USA	1998
DQ811787	*Coronaviridae*	*Alphacoronavirus*	Porcine respiratory coronavirus ISU-1	Pig	USA	2007
NC_018871	*Coronaviridae*	*Alphacoronavirus*	Rousettus bat coronavirus HKU10	Bat	China	2005
DQ811784	*Coronaviridae*	*Betacoronavirus*	Bovine coronavirus DB2	Bovine	USA	1983
NC_030886	*Coronaviridae*	*Betacoronavirus*	Rousettus bat coronavirus	*Rousettus leschenaulti*	China	2014
FJ647223	*Coronaviridae*	*Betacoronavirus*	Murine coronavirus MHV-1	*Mus musculus*	UK	1950

NC_026011	*Coronaviridae*	*Betacoronavirus*	Betacoronavirus HKU24	*Rattus norvegicus*	China	2012
NC_048217	*Coronaviridae*	*Betacoronavirus*	Murine hepatitis virus	Murine	Unknown	Unknown
MT430884	*Coronaviridae*	*Betacoronavirus*	Apodemus peninsulae coronavirus	*Apodemus peninsulae*	China	2019
NC_014470	*Coronaviridae*	*Betacoronavirus*	Bat coronavirus BM48-31/BGR/2008	*Rhinolophus blasii*	Bulgaria	2008
NC_019843	*Coronaviridae*	*Betacoronavirus*	Middle East respiratory syndrome-related coronavirus	*Homo sapiens*	Saudi Arabia	2012
NC_025217	*Coronaviridae*	*Betacoronavirus*	Bat Hp-betacoronavirus/Zhejiang2013	*Hipposideros pratti*	China	2013
NC_009020	*Coronaviridae*	*Betacoronavirus*	Pipistrellus bat coronavirus HKU5	Bat	China	Unknown
NC_012936	*Coronaviridae*	*Betacoronavirus*	Rat coronavirus Parker	Rat	Unknown	Unknown
NC_039207	*Coronaviridae*	*Betacoronavirus*	Betacoronavirus Erinaceus	*Erinaceus europaeus*	Germany	2012
EF065513	*Coronaviridae*	*Betacoronavirus*	Bat coronavirus HKU9-1	Bat	China	2011
JX869059	*Coronaviridae*	*Betacoronavirus*	Human betacoronavirus 2c	*Homo sapiens*	Saudi Arabia	2012
NC_006577	*Coronaviridae*	*Betacoronavirus*	Human coronavirus HKU1	*Homo sapiens*	Unknown	Unknown
NC_006213	*Coronaviridae*	*Betacoronavirus*	Human coronavirus OC43	*Homo sapiens*	USA	Unknown
NC_017083	*Coronaviridae*	*Betacoronavirus*	Rabbit coronavirus HKU14	*Oryctolagus cuniculus*	China	2006
NC_009019	*Coronaviridae*	*Betacoronavirus*	Tylonycteris bat coronavirus HKU4	Bat	China	Unknown
NC_038294	*Coronaviridae*	*Betacoronavirus*	Betacoronavirus England 1	*Homo sapiens*	United Kingdom	2012
MT997203	*Coronaviridae*	*Betacoronavirus*	SARS-CoV-2	*Homo sapiens*	USA	2020
DQ022305	*Coronaviridae*	*Betacoronavirus*	Bat SARS coronavirus HKU3-1	Bat	China	2005
NC_004718	*Coronaviridae*	*Betacoronavirus*	SARS coronavirus Tor2	*Homo sapiens*	Canada	Unknown
NC_016996	*Coronaviridae*	*Deltacoronavirus*	Common-moorhen coronavirus HKU21	Common moorhen	Hong Kong	2007
NC_016994	*Coronaviridae*	*Deltacoronavirus*	Night-heron coronavirus HKU19	Night heron	Hong Kong	2007

NC_039208	*Coronaviridae*	*Deltacoronavirus*	Porcine coronavirus HKU15	Pig	Hong Kong	2010
NC_016995	*Coronaviridae*	*Deltacoronavirus*	Wigeon coronavirus HKU20	Wigeon	Hong Kong	2008
NC_016993	*Coronaviridae*	*Deltacoronavirus*	Magpie-robin coronavirus HKU18	Magpie-robin	Hong Kong	2007
NC_011549	*Coronaviridae*	*Deltacoronavirus*	Thrush coronavirus	Grey-backed thrush	Hong Kong	2007
NC_011550	*Coronaviridae*	*Deltacoronavirus*	Munia coronavirus	White-rumped munia	Hong Kong	2007
NC_016991	*Coronaviridae*	*Deltacoronavirus*	White-eye coronavirus HKU16	White-eye	Hong Kong	2007
NC_011547	*Coronaviridae*	*Deltacoronavirus*	Bulbul coronavirus	Red-whiskered bulbul	Hong Kong	2007
KF793824	*Coronaviridae*	*Gammacoronavirus*	Bottlenose dolphin coronavirus HKU22	Bottlenose dolphin	Hong Kong	2008
MK423877	*Coronaviridae*	*Gammacoronavirus*	Pheasant coronavirus	Pheasant	China	2017
NC_048214	*Coronaviridae*	*Gammacoronavirus*	Duck coronavírus	Duck	China	2014
NC_010800	*Coronaviridae*	*Gammacoronavirus*	Turkey coronavirus	Turkey	Canada	Unknown
NC_010646	*Coronaviridae*	*Gammacoronavirus*	Beluga Whale coronavirus	*Delphinapterus leucas*	Unknown	Unknown
NC_046965	*Coronaviridae*	*Gammacoronavirus*	Canada goose coronavirus	*Branta canadensis*	Canada	2017
NC_001451	*Coronaviridae*	*Gammacoronavirus*	Infectious bronchitis virus	Chicken	Unknown	Unknown
NC_048213	*Coronaviridae*	*Gammacoronavirus*	Infectious bronchitis virus	Chicken	India	2003
AY646283	*Coronaviridae*	*Gammacoronavirus*	Avian infectious bronchitis virus partridge/GD/S14/2003	Partridge	Unknown	2013
NC_048212	*Coronaviridae*	unclassified	Bat coronavirus	*Eidolon helvum*	Cameroon	2013
NC_046955	*Coronaviridae*	unclassified	Shrew coronavirus	*Sorex araneus*	China	2015
NC_034440	*Coronaviridae*	unclassified	Bat coronavirus	*Pipistrellus cf. hesperidus*	Uganda	2013
NC_046954	*Coronaviridae*	unclassified	Rodent coronavirus	*Myodes rufocanus*	China	2014

NC_032496	*Medioniviridae*	*Turrinivirus*	Beihai Nido-like virus 1	Turritella sea snails	China	2014
NC_040361	*Mononiviridae*	*Alphamononivirus*	Planarian secretory cell nidovirus	*Schmidtea mediterranea*	USA	2013
NC_046957	*Nanhypoviridae*	*Sajorinivirus*	Wuhan japanese halfbeak arterivirus	*Hyporhamphus sajori*	China	Unknown
NC_046960	*Nanghoshaviridae*	*Chimshavirus*	Nanhai ghost shark arterivirus	*Chimaera sp.*	China	Unknown
NC_048215	*Roniviridae*	*Okavirus*	Yellow head virus	*Fenneropenaeus chinensis*	China	2012
NC_010306	*Roniviridae*	*Okavirus*	Gill-associated virus	*Penaeus monodon*	Australia	Unknown
NC_046959	*Gresnaviridae*	*Cyclophivirus*	Guangdong greater green snake arterivirus	*Cyclophiops major*	China	Unknown
NC_032492	*Euroniviridae*	*Charybnivirus*	Beihai Nido-like virus 2	*Charybdis crab*	China	2014
NC_032763	*Euroniviridae*	*Charybnivirus*	Wenling nido-like virus 1	Crustacean	China	2014
NC_032490	*Euroniviridae*	*Paguronivirus*	Beihai hermit crab virus 4	Hermit crab	China	2014
NC_038295	*Tobaniviridae*	*Bafinivirus*	Fathead minnow nidovirus	*Pimephales promelas*	USA	1997
NC_008516	*Tobaniviridae*	*Bafinivirus*	White bream virus	*Blicca bjoerkna*	Germany	Unknown
NC_027199	*Tobaniviridae*	*Bostovirus*	Bovine nidovirus TCH5	*Bos taurus*	USA	2013
MT123520	*Tobaniviridae*	*Oncotshavirus*	Chinook salmon bafinivirus	*Carassius auratus*	UK	2018
NC_033700	*Tobaniviridae*	*Infratovirus*	Xinzhou toro-like virus	Snake	China	2014
MG752895	*Tobaniviridae*	*Pregotovirus*	Ball python nidovirus 1	*Python regius*	USA	2016
NC_043474	*Tobaniviridae*	*Pregotovirus*	Shingleback nidovirus 1	*Tiliqua rugosa*	Australia	2015
NC_035465	*Tobaniviridae*	*Pregotovirus*	Morelia viridis nidovirus	*Morelia viridis*	Switzerland	2014
NC_043490	*Tobaniviridae*	*Sectovirus*	Xinzhou nematode virus 6	Snake	China	2014
MG996765	*Tobaniviridae*	*Torovirus*	Berne virus	*Equus caballus*	Switzerland	1972
NC_046956	*Tobaniviridae*	*Torovirus*	Bellinger River virus	*Myuchelys georgesi*	Australia	2015

NC_046962	*Tobaniviridae*	*Torovirus*	Hainan hebius popei torovirus	*Hebius popei*	China	Unknown
NC_046963	*Tobaniviridae*	*Torovirus*	Guangdong red-banded snake torovirus	*Lycodon rufozonatus*	China	Unknown
NC_034976	*Tobaniviridae*	*Torovirus*	Goat torovirus	*Capra hircus*	China	2012
NC_007447	*Tobaniviridae*	*Torovirus*	Breda virus	Bovine	USA	1982
LC483442	*Tobaniviridae*	*Torovirus*	Porcine torovirus	*Sus scrofa*	Japan	2018

Supplementary Table 3: Mean genetic distances calculated for full-length genomes and specific genome regions of mesoniviruses.

	Complete genome	**ORF1a**	**ORF1b**	**ORF2a**	**ORF2b**	**ORF3a**	**ORF3b**	**ORF4**
Mean genetic distance	0.15	0.17	0.10	0.14	0.14	0.13	0.13	0.04

Supplementary Table 4: Pairwise genetic distances between all mosquito mesonivirus sequences for complete genome and RdRp and S protein, for both nt and aa sequences. Excel spreadsheet available online at: https://doi:10.1016/j.virusres.2022.198727

Supplementary Table 5: Pairwise evolutionary distances between all mesonivirus RdRp aa sequences. Excel spreadsheet available online at: https://doi:10.1016/j.virusres.2022.198727

Supplementary Table 6: Assessment of phylogenetic signal of viral sequences by likelihood mapping analysis of the full-length genome sequences of different families in the *Nidovirales* Order.

	Arteroviruses	**Coronaviruses**	**Tobaniviruses**
Totally resolved quartets	99.4%	99.7%	100%
Partially resolved quartets	0.3%	0.2%	0%
Unresolved quartets	0.3%	0.1%	0%

Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Supplementary Table 7: Assessment of temporal signal (root-to-tip) analyses of datasets from different proteins and the full-length genome for mesonivirus nucleotide sequences.

	Mosquito mesoniviruses			**All mesoniviruses**
	Full-legnth genome	**RdRp**	**S**	**RdRp**
Number of sequences	**44**	**44**	**44**	**47**
Date range	36	36	36	37
Slope (rate)	-0.0028	-0.0005	-0.0024	-0.0004
X-Intercept	2176	2500	2140	4924
Correlation coefficient	-0.722	-0.712	-0.74	-0.24
Root-to-tip analysis (r^2)	0.52	0.51	0.55	0.061

Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Supplementary Table 8: Assessment of selective pressure of mesonivirus, based on the analysis of full-length genomes and specific genome regions of mesoniviruses (a) or ORF1ab coding sequences of different families from the *Nidovirales* order (b). The percentage of sites under negative and positive selection were calculated per total number of sites for each genomic region analyzed.

a)

ꝺ	**Complete genome**	**ORF1a**	**ORF1b**	**ORF2**	**ORF3**	**ORF4**
FEL	0.175	0.151	0.113	0.257	0.185	1.16
SLAC	0.211	0.202	0.142	0.280	0.222	1.12
SNAP	0.152	0.127	0.097	0.223	0.146	1.05
Sites under negative selection	**Complete genome**	**ORF1a**	**ORF1b**	**ORF2a**	**ORF3a**	**ORF4**
Number of sites	**5582**	**2240**	**2153**	**855**	**151**	**44**
FEL	2717 (49%)	1147 (51%)	1021 (47%)	335 (39%)	66 (44%)	1 (2%)
SLAC	1735 (31%)	718 (32%)	591 (27%)	201 (24%)	38 (25%)	0
Sites under positive selection	**Complete genome**	**ORF1a**	**ORF1b**	**ORF2a**	**ORF3a**	**ORF4**
Number of sites	**5582**	**2240**	**2153**	**855**	**151**	**44**
FEL	56 (1%)	6 (0.03%)	3 (0.01%)	17 (2%)	0	1 (2%)
SLAC	4 (0.07%)	1 (0.004%)	0	2 (0.2%)	0	0

b)

GD value	**Mesonivirus**	**Arterivirus**	**Coronavirus**	**Tobanivirus**
FEL	0.14	2.02	1.81	1.98
SLAC	0.18	2.12	2.00	2.33
SNAP	0.13	2.61	2.48	2.65
Sites under negative selection	**Mesonivirus**	**Arterivirus**	**Coronavirus**	**Tobanivirus**
Number of sites	**4629**	**1463**	**3264**	**1180**
FEL	2441 (53%)	98 (7%)	287 (9%)	42 (4%)
SLAC	1606 (35%)	2 (0.1%)	15 (0.5%)	0
Sites under positive selection	**Mesonivirus**	**Arterivirus**	**Coronavirus**	**Tobanivirus**
Number of sites	**4629**	**1463**	**3264**	**1180**
FEL	13 (0.2%)	83 (6%)	217 (7%)	27 (2%)
SLAC	1 (0.02%)	34 (2%)	190 (6%)	5 (0.4%)

Readdressing the genetic diversity and taxonomy of the *Mesoniviridae* family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences

Chapter 6. Insect-specific viruses in the *Parvoviridae* family: Genetic lineage characterization and spatiotemporal dynamics of the recently established *Brevihamaparvovirus* genus

Published as:

Morais, P., Trovão, N. S., Abecasis, A., & Parreira, R. (2022). Insect-specific viruses in the Parvoviridae family: genetic lineage characterization and spatiotemporal dynamics of the recently established Brevihamaparvovirus genus. Virus Research, 313 (March), 198728. https://doi.org/10.1016/j.virusres.2022.198728

Abstract

The analysis of the viruses allocated to the recently established *Brevihamaparvovirus* genus (*Parvoviridae* family), which includes all previously known brevidensoviruses, has not yet been carried out on an extensive basis. As a result, no detailed genetic lineage characterization has ever been performed for this group of insect-specific viruses.

Using a wide range of molecular tools, we have explored this taxon by calculating Shannon entropy values, intra- and inter-taxon genetic distances, analyzed sequence polymorphisms, and evaluated selective pressures acting on the viral genome. While the calculated *Brevihamaparvovirus* mutation rates were within the range of those of other parvoviruses, their genomes look to be under strong purifying selection, and are also characterized by low diversity and entropy. Furthermore, even though recombination events are quite common among parvoviruses, no evidence of recombination (either intra or intergenic) was found in the *Brevihamaparvovirus*es sequences analyzed. An extended taxonomic analysis and reevaluation of existing *Brevihamaparvovirus*es sequences, many still unclassified, was performed using cut-off values defining NS1 identity between viral sequences from the *Parvovirus* family. Two existing genetic lineages, Dipteran *Brevihamaparvovirus* 1 and Dipteran *Brevihamaparvovirus* 2, were rearranged and the creation of a new one, Dipteran *Brevihamaparvovirus* 3, was suggested. Finally, despite the uncertainties associated with both the time estimates of the most recent common ancestors, which could span from twenty thousand years before the current era to way earlier (in the last century), and the dispersal routes proposed for *Brevihamaparvovirus*es sequences by phylodynamic reconstruction, the analyses here presented could help define how future studies should be conducted as more isolates continue to be identified in the future, and contribute to eliminating possible analytical biases.

Keywords: *Brevihamaparvovirus*; *Parvoviridae*; Virus taxonomy; Phylogenetic analysis; Spatiotemporal dynamics

1. Introduction

Mosquitoes are important vectors for many pathogenic agents with (re)emerging potential, and many of these correspond to viruses (Gould et al., 2017), some of which pose threats to public health, and may cause epidemics that get considerable worldwide attention (Barzon, 2018). However, over the last decade, in addition to the discovery of many pathogenic viruses in association with hematophagous arthropods, many studies have also brought to light a plethora of so-called insect-specific viruses (ISV). The latter encompass a genetically diverse assemblage of taxonomically distinct viruses, which all share restricted/null replication capacity in vertebrate cells (Calisher and Higgs, 2018; Abudurexiti et al., 2019). Among them stand the viruses of the *Brevihamaparvovirus* genus, which belongs to the *Parvoviridae* family, which was first established in 1975 and groups viruses found in most major vertebrate and invertebrate clades (Cotmore et al., 2014).

Parvoviruses are small (23-28 nm), icosahedral-shaped, non-enveloped viral agents with single-stranded linear DNA (ssDNA) genomes ranging from 4 to 6 kilobases (kb). Two major coding regions determine the expression of non-structural (NS) and structural (VP) proteins, the largest of which (the so-called non-structural protein 1, or NS1) displays a highly conserved helicase superfamily domain with helicase and ATPase activity (Cotmore et al., 2019). Until 2020, parvoviruses were allocated to either the *Densovirinae* (infecting invertebrates) or *Parvovirinae* (infecting vertebrates) subfamilies, with initial subfamily demarcation exclusively supported by the topologies of phylogenetic trees (Muzyczka and Berns, 2001). However, a recent taxonomy revision took into account not only phylogenetic criteria, but also amino acid sequence similarity values calculated from comparisons of the sequences of the NS1 protein or its helicase domain (Pénzes et al., 2020). While high sequence identity for most of the NS1 protein characterized the *Parvovirinae* subfamily, the same did not apply to the *Densovirinae* subfamily. In addition, new densoviruses have also been unexpectedly isolated from vertebrates (Bochow et al., 2015; Yang et al., 2016), adding to the heterogeneous nature of this subfamily, and supporting its division into two distinct ones: the *Densovirinae* and *Hamaparvovirinae*. While hamaparvoviruses share (on average) approximately 30% of NS1 amino acid identity, they only share around 20% of sequence identity when their

helicase domain is compared with that of other parvoviruses. Furthermore, as the insect-specific brevidensoviruses (*Densovirinae* subfamily) shared around 30% of NS1 protein identity with other hamaparvoviruses, they were inserted into the *Hamaparvovirinae* subfamily and the genus renamed *Brevihamaparvovirus* (Pénzes et al., 2020).

Among parvoviruses, the members of the genus *Brevihamaparvovirus* have some of the smallest ssDNA genomes (approximately 4 kb), with three open reading frames encoding two non-structural proteins (NS1, NS2) and a capsid protein (VP) (Bergoin and Tijssen, 2010). While NS1 has been known to be crucial for the initiation of viral DNA replication, NS2 participates in viral egress from the nucleus where viral replication takes place (Chen et al., 2021). The VP gene encodes a capsid protein that is essential for viral entry into host cells and the production of infectious virus (Sánchez-Martínez et al., 2012). Also, the viral genome is characterized by the presence of a unique non-coding terminal palindromic hairpin loop that is required for DNA replication (Afanasiev et al., 1991).

*Brevihamaparvovirus*es have been isolated from various mosquito cell-lines (Afanasiev et al., 1991; Ren et al., 2008) as well as wild mosquitoes, mostly from different species of *Aedes* and *Culex* from Asia (Kittayapong et al., 1999), the Americas (Sadeghi et al., 2018), Europe (Silva et al., 2019) and Africa (Morais et al., 2020), suggesting a widespread distribution. As many of these viruses have been isolated in recent years, we undertook an extensive genetic diversity analysis of this taxon using multiple bioinformatic tools. These included a comprehensive phylogenetic and an attempted spatiotemporal dispersal reconstruction analysis of these ISV.

2. Materials and methods

2.1. Dataset and sequence alignment preparation

The compilation of the different nucleotide (nt) and amino acid (aa) sequence into the datasets used in this work was based on the selection of sequences encoding non-structural protein 1 (NS1), non-structural protein 2 (NS2) and viral protein (VP) of members of the *Brevihamaparvovirus* genus, available at GenBank as of 01/08/2021. These were either directly identified via their respective accession numbers (described in previous publications), or indirectly singled out as a product of sequence similarity

searches using BLASTn. All sequences available to date were downloaded, and additional information including GenBank accession number, host species, geographic origin, and collection date was also obtained. Furthermore, for comparative and phylogenetic purposes, representative datasets containing NS1 nt and aa sequences (the most conserved genomic section) of viruses from other genera in the *Parvoviridae* family were also constructed.

Multiple alignments of viral sequences in each dataset were performed using the iterative G-INS-I method as implemented in MAFFT vs.7 (Katoh and Standley, 2013), followed by their edition using GBlocks (Castresana, 2000). The multiple sequence alignments were systematically verified to ensure the correct alignment of homologous codons using BioEdit 7.0.5 (Hall, 1999). Multiple alignments of NS1 aa sequences from multiple genus in the *Parvoviridae* family were also performed. In this case, the MAFFT iterative L-INS-I option was employed by alignment refinements using the G-INS-I method.

2.2. Assessment of the temporal and phylogenetic signals of Brevihamaparvovirus sequence datasets

The evolutionary information contained in all the sequence datasets compiled (phylogenetic signal) was assessed by Likelihood Mapping (Strimmer and von Haeseler, 1997) using TREE-PUZZLE v5.3 (Schmidt et al., 2002). Datasets for which > 85% of the sequence quartets were totally resolved (randomly selected), were considered of good/high phylogenetic resolving power.

A visual inspection of the degree of genomic sequence divergence accumulation over the sampling time interval (i.e., temporal signal) in all nt datasets was carried out using an exploratory linear regression approach, assuming the topology obtained in a Maximum Likelihood (ML) tree, estimated under an unconstrained clock and the GTR+Γ+I substitution model using IQ-TREE (Trifinopoulos et al., 2016). Root-to-tip divergence values were plotted as a function of sampling time using the TempEst software (Rambaut et al., 2016).

2.3. Genetic diversity analyses

The estimation of genetic distance values (corrected with the Kimura-2P formula) was carried out using MEGAX (Kumar et al., 2018). Heat maps were designed based on pairwise evolutionary distances obtained using the Heatmapper webserver (Babicki et al., 2016). Visualization of genome organization for *Brevihamaparvovirus* (as well as that of selected members of other *Parvoviridae* genera) was executed using Open Reading Frame (ORF) Finder (available in https://www.ncbi.nlm. nih.gov/orffinder/), and the SMART webserver (Letunic and Bork, 2018). The presence of conserved domains in viral protein sequences was investigated using CD-Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). For protein variation analyses, single amino-acid polymorphisms (SAPs) were detected. The indicated amino acid coordinates correspond to those in the *Aedes albopictus* densovirus 2 sequence (accession number X74945).

The analyses of selective pressure acting on individual sites of codon alignments were carried out using the Single Likelihood Ancestor Counting (SLAC) and the Fixed Effects Likelihood (FEL) methods as implemented in Datamonkey (Kosakovsky Pond and Frost, 2005), or the SNAP tool (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP. html), the latter exploring a simpler method for calculation of synonymous and non-synonymous substitutions (Nei and Gojoborit, 1986). Principal coordinate analyses were carried out using PCOORD (http://www.hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html). Additionally, possible recombination events were investigated using the Recombination Detection Program 4 (RDP4) software (Martin et al., 2015).

2.4. Phylogenetic analyses using maximum likelihood and Bayesian approaches

To assess the relationship between the isolates belonging to the *Brevihamaparvovirus* genus and the remaining genera in the *Parvoviridae* family, phylogenetic reconstructions were carried out using NS1 aa sequences and the ML optimization criterion assuming the Whelan and Goldman (WAG) model, as defined by IQ-TREE (Trifinopoulos et al., 2016). The stability of the obtained tree topologies was assessed by bootstrapping based on 1000 re-samplings of the original sequence data. Phylogenetic reconstructions (ML) using *Brevihamaparvovirus* ORF-specific nt and aa datasets were performed using the

GTR+Γ+I and WAG model, respectively, as suggested by IQ-TREE. Once again, the stability of the obtained tree topologies was assessed by bootstrapping with 1000 re-samplings of the original sequence data. All phylogenetic reconstructions were carried out assuming a relaxed uncorrelated lognormal molecular clock model (Ho et al., 2005) as indicated by the ML Clock Test implemented in MEGA X, allowing for the accommodation of among-lineage rate variation.

Time-calibrated phylogenetic and phylogeographic histories were obtained using a Bayesian statistical framework, as implemented in the BEAST v1.10 software package (Suchard et al., 2018), and using the GTR+Γ+I model. To investigate the sensitivity of the estimate for the time to the Most Recent Common Ancestor (tMRCA) concerning the coalescent priors used, the performance of constant, exponential, logistic, and expansion parametric population demographic growth priors (Drummond et al., 2003; Griffiths and Tavaré, 1994) was tested against that of nonparametric ones, including the Bayesian Gaussian Markov Random Field (GMRF) Skyride (Minin et al., 2008), Skygrid (Gill et al., 2013) and Skyline (Drummond et al., 2005). This preliminary comparative analysis was carried out using all the VP dataset sequences available. Bayes factor (BF) support for predictors was calculated using marginal likelihood estimates (MLE) (inferred using Path Sampling (PS) and Stepping-Stone (SS) approaches) for each candidate model, and then comparing the ratio of the marginal likelihood estimates for the set of candidate models being compared.

A minimum number of two, and up to a maximum of twenty, independent Markov chain Monte-Carlo (MCMC) runs were performed using BEAST v1.10 until 1-3 × 108 states were sampled, with at least 10% of which being discarded as burn-in. The length of the MCMC analyses was defined as a function of chain convergence which was followed using the Tracer software v1.7.1 (http://beast.bio.ed.ac.uk/tracer). The latter was also used to check for adequate effective sample size (ESS) higher than 200 after the removal of burn-in. The trees were logged on every 10,000th MCMC step, and the trees distribution was summarized using the TreeAnnotator software v1.8.3 as a maximum clade credibility (MCC) tree, with median heights as the node heights in the tree. The FigTree v1.4.2 software was used to visualize the phylogenetic trees (htt p://tree.bio.ed.ac.uk/software/figtree/).

2.5. Continuous phylogeography

The geographic spread of *Brevihamaparvovirus*es in continuous space was studied using a phylogenetic Brownian diffusion approach that models the change in geographic coordinates (latitude and longitude) along each branch in the phylogenetic reconstruction (Lemey et al., 2010). As an alternative to the latter, relaxed random walk (RRW) extensions that model branch-specific variation in dispersal rates similar to uncorrelated relaxed clock approaches was also used (Drummond et al., 2006). The assessment of BF support for the diffusion priors was calculated using MLE as described above for the coalescent demographic priors.

The spatiotemporal reconstruction of the spread *Brevihamaparvovirus* was visualized on the Spatial Phylogenetic Reconstruction of Evolutionary Dynamics software (SpreaD3; Bielejec et al., 2016), using a custom-made geoJSON world map (https://geojson-maps.ash.ms/).

3. Results

3.1. Comparative genomic coding architecture and genetic diversity analyses

Public genomic database mining allowed the creation of three datasets containing *Brevihamaparvovirus* (BHP) nt and aa sequences. These included 60 NS1 sequences, 31 NS2 sequences, and 40 VP sequences (additional information available on Supplementary Table 1). Most sequences (~90%) were originally identified in association with specimens of either *Culex sp.* or *Aedes sp.* mosquitoes, with the remaining ones being amplified from *Anopheles*, *Culiseta*, *Armigeres*, and *Haemagogus*. Five of the BHP sequences were obtained from C6/36 cell cultures and two from chronically infected cell lines (Aag2 and SuaB5). Additionally, for phylogenetic and other comparative analyses, NS1 aa sequences were also compiled for viruses from each genus/species in all *Parvoviridae* subfamilies (Supplementary Table 2).

Brevihamaparvovirus ORF organization, as seen in Fig. 1, is almost identically shared between all BHP, with two distinct regions coding the non-structural proteins (NS1 and NS2) and one viral structural protein (VP). However, the genome of one BHP (accession

number MH188047), isolated in 2016 from a *Culex* mosquito, displayed two ORFs that encoded structural proteins (Fig. 1). This genetic organization recalls that of typical *Ambidensovirus* (*Densovirinae* subfamily) sequences (Supplementary Fig. 1). The composition of ORF-coding sequences looked similar between different genera in the *Hamaparvovirinae* subfamily, with the only noticeable exception found among the members of the *Ichthamaparvovirus* genus, for which an NS2 coding sequence could not be identified. On the contrary, ORF organization inside the *Densovirinae* and *Parvovirinae* subfamilies was found to be quite disparate. An alternative ORF-coding sequence, encoding the so-called assembly-activating protein (AAP), which promotes capsid assembly (Earley et al., 2017), seemed to exist only in the genome of the members of the *Dependoparvovirus* genus.

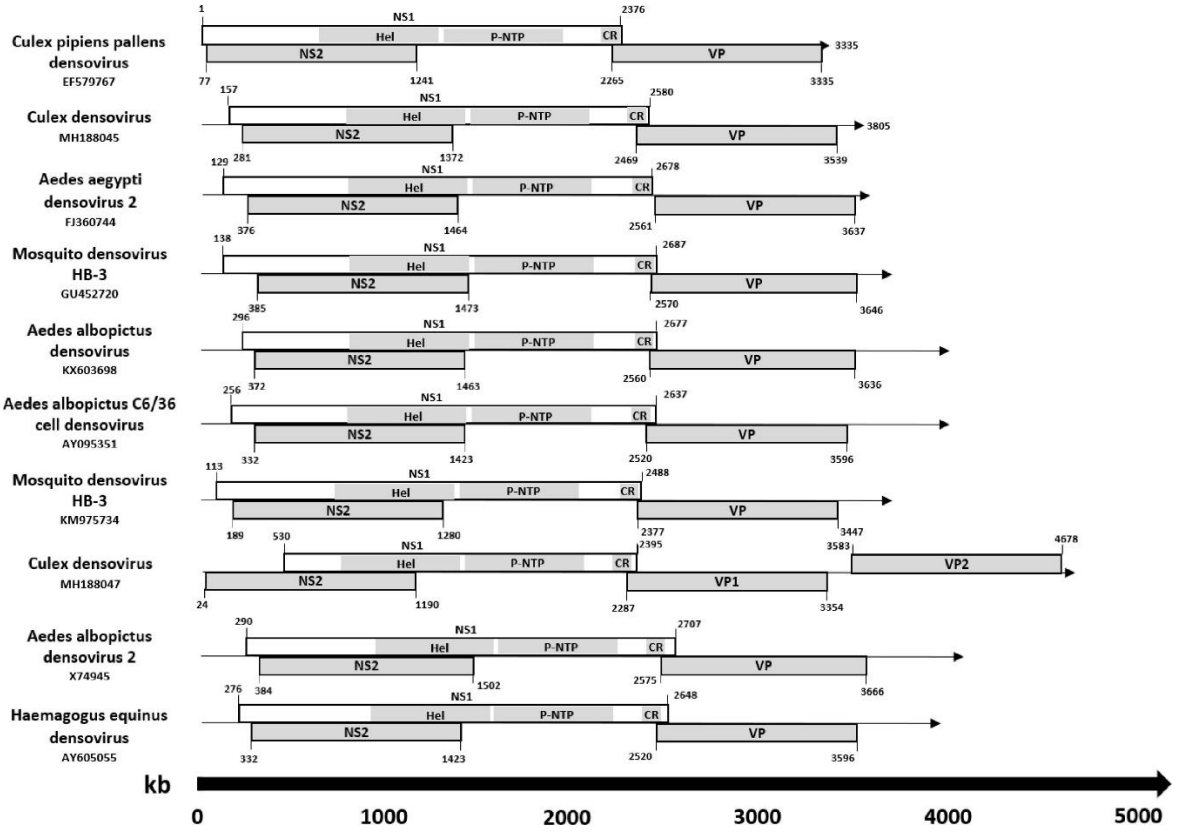


Fig. 1: Schematic representation of nucleotide sequences for ten different Brevihamaparvovirus, with different ORFs identified; NS – non-structural protein; VP – viral protein; Hel – Helicase; P-ATP – P-loop NTPase; CR – coiled region.

Overall, comparison of all individual BHP ORF-coding sequences disclosed low mean genetic distances, with the lowest value associated with the NS1 protein (0.053), followed by NS2 (0.064) and VP (0.077). Pairwise evolutionary distances (PEDs) were calculated between all BHP sequences and analyzed using a heatmap (Supplementary Fig. 2). PEDs analysis did provide insights into possible segregation of three different groups of sequences inside the *Brevihamaparvovirus* genus, with those inside each group sharing low PEDs values, while slightly higher values were observed when sequence comparisons extended towards those from other viral groups. To compare genetic distance values between members of the different *Parvoviridae* subfamilies, overall mean genetic distances were calculated individually (for each subfamily) for the most conserved ORF-coding region (NS1), using datasets containing sequences from each genus/species. The inclusion of a more divergent group of NS1-coding sequences into a single dataset naturally raised the average genetic distance values of the *Hamaparvovirinae, Densovirinae* and *Parvovirinae* subfamilies to 0.499, 0.502 and 0.503, respectively, with no apparent significant difference between all values.

Shannon entropy is a quantitative measure of uncertainty in a dataset of nucleotide or amino acid sequences, and it may be considered as a measure of variation in DNA and protein sequence alignments for assessment of genetic diversity in a cross-sectional sense. When applied to the analysis of BHP sequences, Shannon entropy calculations showed low values for all BHP ORF-coding sequences. NS1-coding region did show slightly lower entropy values in most amino acids (Supplementary Fig. 3A), which should be expected due to its key role in viral DNA replication. In addition, no substantial differences in entropy values were found between different subfamilies in the *Parvoviridae* families when evaluating NS1 aa sequences entropy (Supplementary Fig. 3B).

3.2. Phylogenetic signal, selective pressure, impact of genetic recombination, and sequence divergence accumulation throughout time

In order to assess the extent to which selective pressure and/or intra/ intergenic recombination events could impact phylogenetic reconstructions, both were metrics

evaluated using specific bioinformatic tools. No evidence of either intra or intergenic recombination events were detected for either full-length genomes or any of the genes analyzed, using all detection methods on the RDP4 software. Estimation of omega (ω) values (corresponding to the ratio of non-synonymous to synonymous substitutions) was performed for BHP using three different methods (SLAC, FEL, and SNAP). These analyses were carried out for all ORF-coding regions (Supplementary Table 3), and the results obtained indicated that the whole genome seems to be under purifying selection, as deduced by overall low ω values, especially in the case of the VP-coding region (*p*-value < 0.05). Site-specific selection analysis also revealed little to no evidence of positively selected sites. Only 2 codons in the NS2 gene were identified as evolving under diversifying selection, and even so, this observation was only supported by one analysis methods used (FEL).

Since no evidence of recombination events or positive selective pressure acting on the BHP genome were detected (which could compromise phylogenetic reconstruction), the phylogenetic signals of each nt dataset were evaluated using likelihood mapping (Table 1). The obtained results showed high phylogenetic resolving power for both the NS1 and VP sequence datasets, with 88.1% and 85.5%, of totally resolved randomly selected 10,000 quartet replicates respectively. The NS2 gene showed slightly lower phylogenetic resolving power with 77.0% of totally resolved sequence quartets. These results indicated that phylogenetic reconstructions based on the analysis of alignments of any viral ORFs (with the possible exception of NS2), would produce unambiguous phylogenetic trees.

Table 1: Phylogenetic signal (as assessed by likelihood mapping) and root-to-tip (sequence divergence as a function of time) of brevihamaparvovirus sequences using datasets of all ORF-coding sequences.

Likelihood Mapping	**NS1**	**NS2**	**VP**
Totally resolved quartets	88.1%	77.0%	85.5%
Partially resolved quartets	3.5%	0.7%	1.8%
Unresolved quartets	8.4%	22.3%	12.7%
Root-to-tip analysis (r^2)	0.028	0.220	0.450

To assess the extent to which all BHP sequence datasets contained detectable signals indicating expanding sequence divergence throughout time, a standard linear regression exploration of root-to-tip genetic distances as a function of sampling time was performed. Only the NS1-coding region did not reveal clear evidence for a substantial temporal signal (Table 1), even after the removal of outlier sequences that could have a negative impact on temporal signal assessment. Nevertheless, while for both the NS2 and VP sequence datasets, a substantial temporal signal was found, we selected the VP gene as the prime candidate for continuous phylogeography analysis (see below) as it also displayed the highest phylogenetic signal. However, both the very narrow temporal date sampling interval (with an average date-range of 20 years) and the *a priori* unknown average rates of evolutionary change for *Brevihamaparvovirus*es, could influence temporal signal assessment. As far as the latter was concerned, nucleotide substitution rates were estimated using the sequences of the BHP VP gene, while assuming a relaxed molecular clock model (Drummond et al., 2006). This was supported by ML test of the molecular clock hypothesis, which systematically rejected the null hypothesis of equal nucleotide substitution rates along the branches of the trees (Supplementary Table 4A). Substitution rate values varied depending on the coalescent priors used and ranged from 1.16×10^{-3} to 2.24×10^{-4} substitutions per site/per year.

3.3. Phylogenetic analyses

Previous reports have stated that NS1 proteins of viruses belonging to the same genus share between 35-40% of amino acid sequence identity with a minimum shared query cover of 80% (between any two members being compared), while simultaneously clustering as a robust monophyletic lineage (Pénzes et al., 2020). Accordingly, a phylogenetic reconstruction of the evolutionary relationships within the *Parvoviridae* family (Supplementary Fig. 4) and the genetic distance values indicated in Supplementary Table 5 clearly show the singularity of the *Brevihamaparvovirus* genus. While the analysis of NS1 sequences suggested that *Brevihamaparvovirus*es share common ancestry with the members of *Parvovirinae* subfamily, this shared ancestry was not supported by bootstrap analysis.

Considering the above mentioned, (i) high phylogenetic signal of NS1 and VP sequence datasets, and (ii) the absence of evidence for intra/ intergenic recombination, (iii) or of positive selection acting as a driver of virus evolution, the evolutionary relationships between only BHP were investigated using ML phylogenetic tree reconstruction (Supplementary Fig. 5). No substantial differences were found between the NS1 and VP ML trees, and in both the BHP sequences were segregated into three distinct monophyletic clades. Furthermore, when the previously defined minimum of 85% of identity (based on the analysis of NS1 aa sequences) was considered to unite *Brevihamaparvovirus*es as members of the same species (Pénzes et al., 2020), the different BHP genetic lineages did seem to correspond to distinct viral species, consistent with the NS1 tree topology. However, not only have many new, and therefore unclassified, sequences been described recently from the Americas (Sadeghi et al., 2018), Asia (Fu et al., 2017), Europe, and Africa (Morais et al., 2020; Silva et al., 2019), a dissent was observed between the NS1 tree topology and the currently accepted taxonomic assignments (Fig. 2A). For these reasons, Fig. 2B indicates a suggested correction of the BHP genetic lineage assignment, confirming the previously defined Dipteran *Brevihamaparvovirus* (DB) 1 and DB2 genetic lineages, and suggesting the establishment of a third one, designated DB3. The existence of the DB1-3 genetic sublineages was also supported by PCOORD analysis (Fig. 2C), as well as by the NS1 genetic diversity grouping indicated by heatmaps (Supplementary Fig. 2).

Multiple single amino-acid polymorphisms were identified when NS1 aa variation was analyzed taking into account sequence comparisons between different *Brevihamaparvovirus* branches, further supporting the identified viral sublineages (Fig. 2C). The analysis of lineage-specific SAPs indicated four of them to be exclusively found on DB2 (479D, 487S, 524A and 546H) and DB3 (74H, 496D, 539K and 586S) sequences, while polymorphisms 481R and 522K characterized the DB1 lineage.

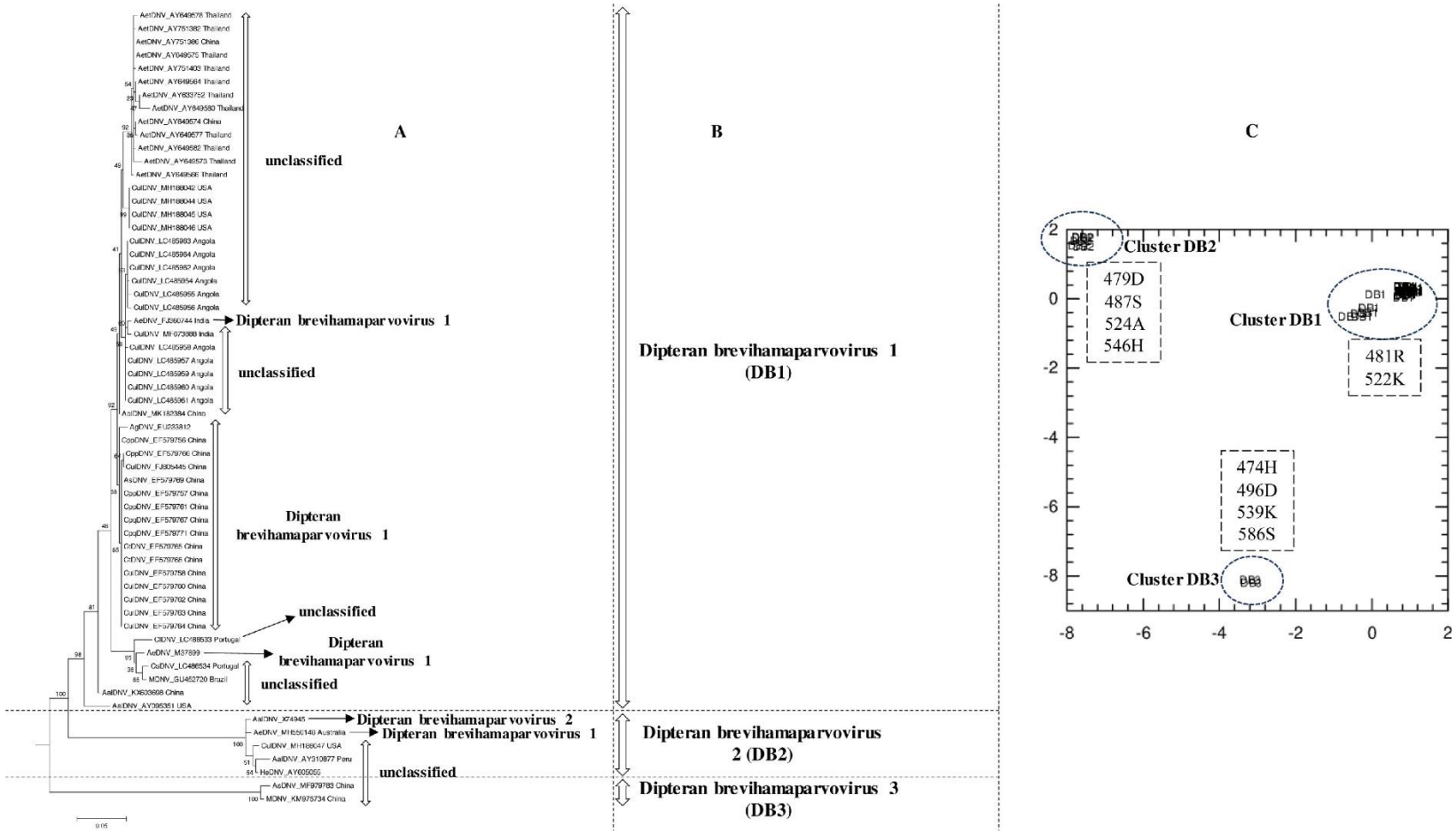


Fig. 2: Maximum likelihood tree of NS1 Brevihamaparvovirus sequences juxtaposed to (A) the current ICTV taxonomy information and (B) a classification scheme of parvoviruses combining the tree topology and the species-defining NS1 identity percentages (as previously defined by Pénzes et al. (2020); (C) Principal coordinate analysis carried out for NS1 coding sequences, where each sequence is identified by the major branch they belong to (as shown in Fig. 1B). Single amino-acid polymorphisms exclusive to each identifiable group are indicated in association with the DB1-3 clusters. The indicated positions relate to those of the Aedes albopictus densovirus 2 sequence (accession number X74945).

3.4. Continuous phylogeography

In an attempt to infer the population dynamics of BHP sequence dispersal through space and time, using a dataset with high phylogenetic resolving power and reasonable temporal signal (VP; r2=0.45), we first tested the performance of parametric demographic priors against that of non-parametric ones. The obtained results clearly showed that the non-parametric priors consistently performed better than the parametric ones, indicated by both Bayes factor (Supplementary Table 4A) and effective sample size (ESS) values, which were consistently higher for non-parametric priors. The obtained results also pointed towards the Bayesian Skyline as the coalescent prior of choice, as judged by marginal likelihood and ESS values (consistently higher than 200). A comparative assessment of the performance of a strict Brownian vs. several RRW diffusion models for BHP was also performed, allowing us to evaluate what would be the best geographic diffusion model to be used for spatiotemporal dispersal analysis. The obtained MLE values (shown in Supplementary Table 4B) suggested that a Gamma-RRW prior was the one best fit to explain its dispersal dynamics.

Results for the spatiotemporal analysis were summarized as a MCC phylogenetic tree (Supplementary Fig. 6), as well as projected into maps using the SpreaD3 software (Fig. 3). High Highest Probability Density (HPD) intervals for MRCA ages were estimated for almost all nodes, especially for the MRCA for all BHP sequences. While these high HPD intervals appear to dissipate as the estimates moved towards the more recent nodes, the analysis of root age dates should be interpreted with caution. Although it seems clear that two viral lineages have diverged well in the past into two distinct clades, evaluation of dispersion routes between BHP's oldest and also the most recent ancestors were not firmly supported by our analyses. With the available data, our analysis suggests the possibility of an expansion of one of the viral lineages into two distinct ones dating more than two thousand years ago (95% HDP: (-10263)-2014) in both eastern and western directions (Asia and North America). The other expansion event, for which a Middle Eastern (possibly Iranian) geographical origin was suggested (though not statistically supported), seemed to have split into two different routes in the 1700s. More recent years marked the expansion of one of the possible Indian clades into Africa (Angola) in the early 2010s (95% HPD: 1990- 2016, and strongly supported with a posterior probability of 1), while the other clade seemed to have expanded in two totally different directions.

These included movements towards Asia, starting in the early 1870 (95% HPD: 1068-1995) as well as North America, starting as late as in 2014 (95% HPD: 1995-2016). Both dispersal routes were strongly supported by location posterior probability values (of 0.8 and 1, respectively).

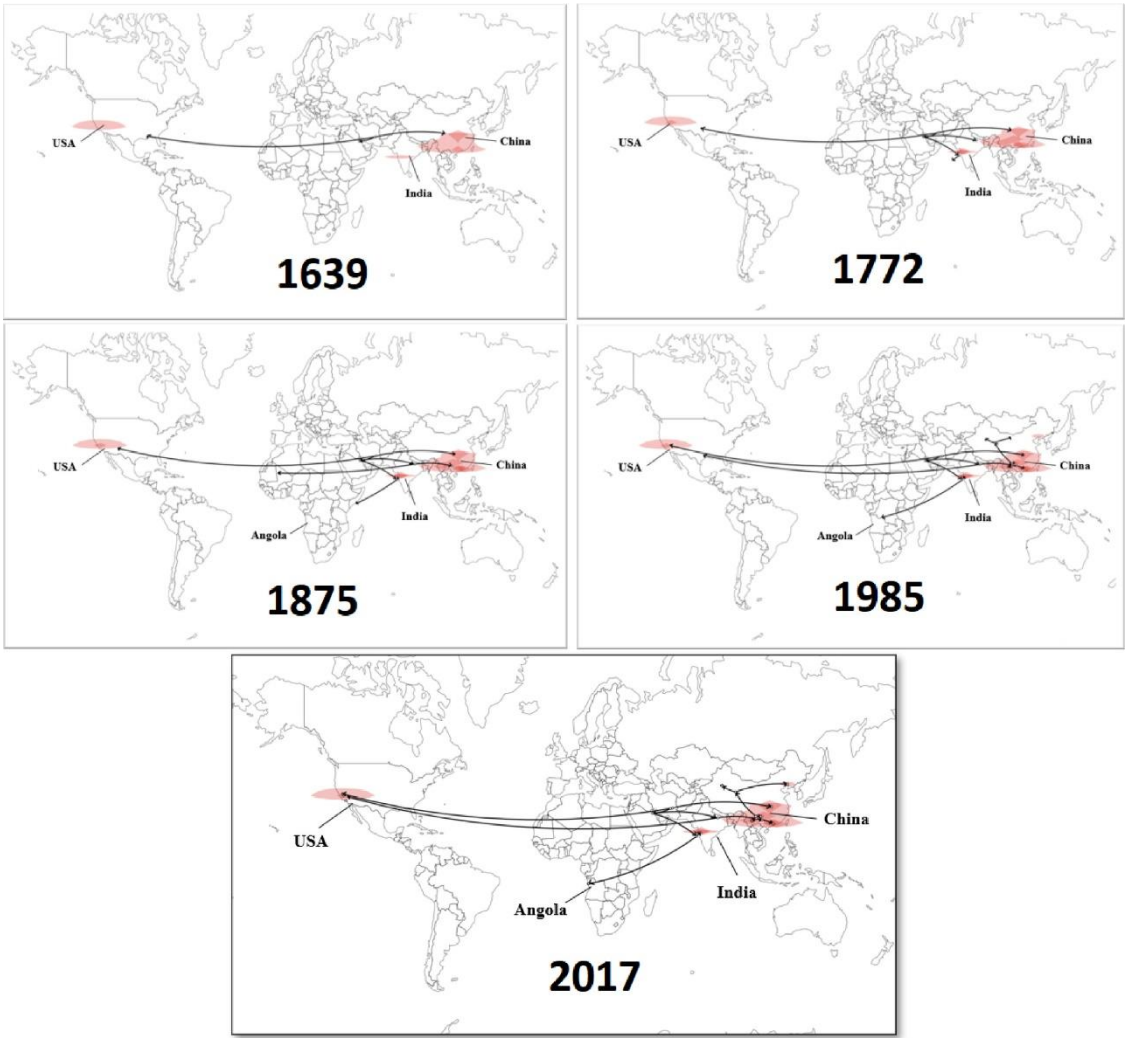


Fig. 3: Spatiotemporal reconstruction of Brevihamaparvovirus spread visualized on SpreaD3 software, based on the MCC tree represented in Supplementary Fig. 6.

4. Discussion

Unlike the case of most genera in the *Parvoviridae* family that join viruses identified in association with both vertebrate and invertebrate species (Pénzes et al., 2020), the members of the *Brevihamaparvovirus* genus have only been, up to the present day, found in mosquitoes. To what extent is this association with Diptera absolute is still open to debate. In fact, the analysis of other ISVs has shown that in specific cases they eventually bypass host range restrictions imposed by certain host cells (Junglen et al., 2017), thus expanding their host range (Morais et al., 2021). These observations open the possibility that among the large diversity of viruses associated with viral *taxa* whose members are supposedly restricted to replicate in insect cells, some may acquire the capacity to adapt to a larger collection of hosts. Such a case has already been described in the *Parvoviridae* family, when sequences of members of the *Ambidensovirus* genus (*Densovirinae* subfamily), mostly associated with insect hosts until recently, have been recently detected in vertebrate hosts such as ducks (accession number MW306771) and cranes (accession number MW046535).

As it has been previously considered for other ISVs (Goenaga et al., 2020), *Brevihamaparvovirus*es are not only widespread in a variety of wild mosquito species, but have also been found to interfere *in vitro* with the replication of *bona fide* arboviruses, such as reducing the severity of the cytopathic effects induced by dengue virus infection in C6/36 cells persistently infected with Aedes albopictus *Brevihamaparvovirus* (Burivong et al., 2004). These observations highlight the potential use of certain arboviruses as biological agents to interfere with vector competence. These viruses seem to be able to integrate their genomes into that of infected cells, therefore they can also be exploited as vehicles for stable expression of heterologous proteins in insect cells (Ohlund et al., 2019). All these facts justify a more detailed approach to gather information on this specific genus in the *Parvoviridae* family, where extensive genetic research has been scarce. In fact, studies addressing the analysis of members of the *Brevihamaparvovirus* genus (formerly designated as *Brevidensovirus*) have been almost non-existent, with only sporadic reports on the detection of a new viral genome, the analysis of its genetic structure (Chen et al., 2004), or what phenotypic effects are associated with their replication on specific cell lines (Paterson et al., 2005). Broader genetic studies have

focused exclusively on either the taxonomy revision of the *Parvoviridae* family as a whole (Pénzes et al., 2020) or the characterization of the phylogenetic relationships of its members (Cotmore et al., 2019), with sporadic reports of estimation of nucleotide substitution rates and selective pressure analysis for some parvovirus (Stamenkovic et al., 2016). However, to this date, the members of the *Brevihamaparvovirus* genus have not been analyzed in detail, including assessments of genetic diversity, selective pressure, Shannon entropy, or spatiotemporal dynamics. In this regard, this study provides new insights into the genetic characteristics of BHP, as well as what evolutionary events may have contributed to their dissemination, and what technical aspects limit our ability to describe it precisely and with detail.

Our genetic analyses were based on the assembly of multiple datasets of sequences from three different genomic regions (NS1, NS2 and VP). The NS proteins share the majority of conserved domains and a coiled region essential for DNA replication (Bergoin and Tijssen, 2010). The regions encoding NS1 and NS2 overlap, but each protein is encoded from a distinct reading frame after alternative splicing (Chen et al., 2021). While two of the regions displayed high phylogenetic signal (NS1 and VP), only one revealed a strong temporal signal (VP-coding region) required for spatiotemporal dispersal analysis. The genomic regions with the lowest and highest overall mean genetic diversity were the NS1 and VP regions, respectively. These results confirm previous observations made in association with the study of other parvoviruses (Kapoor et al., 2010; Lu et al., 2020), for which mean genetic distance values were consistently higher for coding regions of structural proteins when comparing them to non-structural proteins. Recombination events, which seem to commonly affect the evolution parvoviruses of equine (Lu et al., 2020), geese and Muscovy ducks (Shen et al., 2015), or even humans (Khamrin et al., 2013) seem to affect the VP region in particular. However, no recombination events seemed to have affected the evolution of the analysed BHP. Since the generation of recombinant genomes requires co-/superinfection events to occur, whether the apparent lack of impact of recombination on viral evolution is a direct consequence of restricted replication only in insects (mosquitoes in particular) is open to debate.

Positive selection pressure is among the factors that affect virus evolution. In the context of this study, its analysis indicated that most of the BHP genome evolves under strong purifying selection, as seen by the accumulation of a surplus of synonymous substitutions

relative to the non-synonymous substitutions. Only two codons in NS2 may be under diversifying selection, but this observation was not confirmed by all method of analysis used. Furthermore, analysis of Shannon entropy, used as a measure of variation in DNA/protein sequence alignments, revealed low values for all genomic regions. Similar results have been found in canine parvovirus and human parvovirus (Shackelton et al., 2005; Stamenkovic et al. 2016). The biological relevance of these observations should, however, be considered with caution as they may vary significantly depending on the number of BHP sequences available (Añez et al., 2011), and a considerable number of these have been described recently (Morais et al., 2020; Silva et al., 2019). Further research regarding the clarification of what may be the mechanisms of natural selection affecting BHP is important. Indeed, reports of positive selective pressure acting on selected parvovirus genomic sites (especially in capsid protein coding regions) have been strongly connected to their ability to adsorb to new host cells (Hueffer et al., 2003a), allowing possible early detection of future BHP host-switching. Furthermore, whereas the calculated BHP substitution rates are similar to those of other parvoviruses (Shackelton et al., 2005, calculated for canine parvoviruses, ranging from 2.7×10^{-3} to 9.4×10^{-5} substitutions per site/per year; Stamenkovic et al., 2016, calculated for human parvovirus B19, ranging from 1.03×10^{-4} to 2.32×10^{-4} substitutions per site/per year), they are high due to the small size of the viral genome and its high replication turnover (Koppelman et al., 2007), which could influence future processes of natural selections by mutation fixation.

Considering (i) the cut-off value of NS1 sequence identity that seems suitable to define independently evolving genetic lineages for BHP, (ii) the congruence between the topologies of phylogenetic trees, (iii) the recently described BHP sequences that remained unclassified, (iv) with one of them (MH550148) being previously misclassified, our analysis did confirm the establishment of two species (Dipteran *Brevihamaparvovirus* 1 and 2) and suggests that a new one (Dipteran *Brevihamaparvovirus* 3) should be considered. This suggestion was also supported by PCOORD analysis as well as by the identification of NS1 species-specific aa polymorphisms.

Addressing phylodynamic analysis in a BHP genetic characterization study could provide relevant information regarding the estimation of the time and geographic origin of most recent common ancestor (MRCA) of all BHP sequences analysed, as well as what may

have been the routes these viruses explore to spread through space and time. However, a prior assessment of genetic information contained in viral sequences is vital, since not only weak temporal signals could negatively impact the calculations of the mean MRCA time estimates (or tMRCA; Trovao et al., 2015), but the use of adequate candidate models in phylogeographic analyses is of paramount importance. Except for NS2, both NS1 and VP coding sequences showed sufficient temporal signals, which seem to be a common observation when studying parvoviruses (Stamenkovic et al. 2016). However, unlike observations stated in previous reports where the spatiotemporal dynamics of parvovirus was investigated considering constant or logistic population size priors, in thus study we formally demonstrate that non-parametric coalescent priors often perform better than parametric ones (Morais et al., 2021). Therefore, using adequately selected coalescent and demographic dispersal priors, our results suggested a scenario where the MRCA of the BHP under analysis may have emerged up to twenty thousand years before the current era. However, given the large 95% HDP intervals estimated for internal nodes, the proposed ages for mean tMRCA, despite giving an indication of the times of divergence, are not accurate and should be interpreted with caution. The analysis of BHP VP sequences revealed a tree root in the Middle East (with low statistical support) from where two viral lineages diverged. Despite the wide host-range and transmission routes different parvoviruses have explored to ensure their natural maintenance, in reports regarding the analysis of canine parvovirus (Giraldo-Ramirez et al., 2020), the estimated average tMRCA was 1979 (with 95% HPD range of 38–44 years), and the spread of these viruses seems to have happened quite recently. Therefore, given the very wide 95% HPD intervals associated with the older branches of the BHP phylogeography tree, the early BHP expansion events could have occurred considerably later than our analysis suggested. However, this discrepancy might be due to the use of inappropriate priors during the phylodynamic reconstruction. Multiple factors, such as massive tourism and global commercial trading, and the limited success of most vector-control programs could have a strong impact in the recent expansion of BHP. Despite its limitations, this study did provide information that could help define how future studies should be conducted, as more BHP sequences are identified.

Few reports delving into the evolutionary events that shaped the evolution of parvoviruses have been carried out to this day and, not surprisingly, these usually address pathogenic

viruses. To date, none of these studies had specifically focused on the members of the *Brevihamaparvovirus* genus, which stand unique among parvoviruses thanks to their host-restriction. However, host-switching looks to be quite common in parvovirus (Hueffer et al., 2003b), so as new BHP sequences are identified in the near future, new research is crucial in order to identify how and when putative host-switching events might have eventually occurred.

Supporting information

Supplementary Fig. 1: Schematic representation of nucleotide sequences from each genus in the *Parvoviridae* family with different ORFs identified. Not representative of the size of each ORF, only their organization and sequence. NS – non-structural protein; VP – viral protein; NP – nucleoprotein; AAP – assembly activating protein.

Supplementary Fig. 2: Heat map representing inter-sequence genetic diversity of *Brevihamaparvovirus*. Representative tree obtained on IQ-TREE (maximum likelihood, GTR+Γ+I model) based on NS1 nt sequences (reported in Supplementary Table 1), and Z-Scores estimated based on pairwise evolutionary distances using MegaX.

Supplementary Fig. 3: (A) Entropy on the basis of the Shannon function (Shannon entropy-one) for different ORF-coding sequences of *Brevihamaparvovirus*; (B) Entropy on the basis of Shannon function (Shannon entropy-one) for NS1-coding sequences of different subfamilies in the *Parvoviridae* family.

Supplementary Fig. 4: NS1 maximum likelihood phylogenetic tree of several parvovirus genera and subfamilies, estimated under a WAG substitution model using IQ-TREE (phylogeny test with 1000 bootstrap replications). Isolates are shown in Supplementary Table 2.

Supplementary Fig. 5: Maximum likelihood tree of *Brevihamaparvovirus* NS1 and VP nucleotide sequences, estimated under a GTR+Γ+I substitution model using IQ-TREE (phylogeny test with 1000 bootstrap replications). The different genetic lineages (DB1-3) are indicated.

Supplementary Fig. 6: Continuous phylogeographic analysis of *Brevihamaparvovirus* VP coding sequence. At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being

displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles, while the decimals associated with certain nodes indicate the inferred location PP.

CRediT authorship contribution statement

Paulo Morais: Data curation, Formal analysis, Visualization. **Nídia S. Trovão:** Writing – review & editing. **Ana B. Abecasis:** Writing – review & editing. **Ricardo Parreira:** Conceptualization, Supervision, Writing – review & editing. **Declaration of Competing Interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work received financial support from the Global Health and Tropical Medicine Center, which is funded through FCT contract UID/ Multi/04413/2013. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2022.198728.

References

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S.V., Avsic-Zupanc, T., Ballinger, M.J., Kuhn, J.H., 2019. Taxonomy of the order bunyavirales: update 2019. Arch. Virol. 164 (7), 1949–1965. https://doi.org/10.1007/s00705-019-04253-6.

Afanasiev, B.N., Galyov, E.E., Buchatsky, L.P., Kozlov, Y.V., 1991. Nucleotide sequence and genornic organization of aedes densonucleosis virus. Virology 185 (1), 323–336. https://doi.org/10.1016/0042-6822(91)90780-F.

Añez, G., Morales-Betoulle, M.E., Rios, M., 2011. Circulation of different lineages of dengue virus type 2 in Central America, their evolutionary time-scale and selection pressure analysis. PLoS One 6 (11). https://doi.org/10.1371/journal.pone.0027459.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S., 2016. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 44 (W1), W147–W153. Barzon, L., 2018. Ongoing and emerging arbovirus threats in Europe. J. Clin. Virol. 107, 38–47. https://doi.org/10.1016/j.jcv.2018.08.007.

Bergoin M. and Tijssen P. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses, p. 57–90. *In* Asgari N. and Johnson K. N. (ed.), *Insect virology*. Horizon Scientific Press, Norwich, United Kingdom.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2016. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol. Biol. Evol. 33, 2167–2169. https://doi.org/10.1093/molbev/msw082.

Bochow, S., Condon, K., Elliman, J., Owens, L., 2015. First complete genome of an ambidensovirus; cherax quadricarinatus densovirus, from freshwater crayfish cherax quadricarinatus. Mar. Geonomics 24, 305–312. https://doi.org/10.1016/j.margen.2015.07.009.

Burivong, P., Pattanakitsakul, S.N., Thongrungkiat, S., Malasit, P., Flegel, T.W., 2004. Markedly reduced severity of dengue virus infection in mosquito cell cultures persistently infected with aedes albopictus densovirus (AalDNV). Virology 329 (2), 261–269. https://doi.org/10.1016/j.virol.2004.08.032.

Calisher, C.H., Higgs, S., 2018. The discovery of arthropod-specific viruses in hematophagous arthropods: an open door to understanding the mechanisms of arbovirus and arthropod evolution? Annu. Rev. Entomol. 63, 87–103. https://doi.org/10.1146/annurev-ento-020117-043033.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Chen, S., Cheng, L., Zhang, Q., Lin, W., Lu, X., Brannan, J., Zhang, J., 2004. Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected aedes albopictus C6/36 cell line. Virology 318 (1), 123–133. https://doi.org/10.1016/j.virol.2003.09.013.

Chen, S., Miao, B., Chen, N., Chen, C., Shao, T., Zhang, X., Tong, D., 2021. SYNCRIP facilitates porcine parvovirus viral DNA replication through the alternative splicing of NS1 mRNA to promote NS2 mRNA formation. Vet. Res. 52 (1), 1–15. https://doi.org/10.1186/s13567-021-00938-6.

Cotmore, S.F., Agbandje-McKenna, M., Chiorini, J.A., Mukha, D.V., Pintel, D.J., Qiu, J., Davison, A.J., 2014. The family *Parvoviridae*. Arch. Virol 159 (5), 1239–1247. https://doi.org/10.1007/s00705-013-1914-1.

Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., Chiorini, J.A., Eis-Hubinger, A.M., Hughes, J., Harrach, B., 2019. ICTV virus taxonomy profile: *Parvoviridae*. J. Gen. Virol. 100 (3), 367–368. https://doi.org/10.1099/jgv.0.001212.

Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. Trends Ecol. Evol. 18, 481–488.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192. https://doi.org/10.1093/molbev/msi103.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710. https://doi.org/10.1371/journal.pbio.0040088.

Earley, L.F., Powers, J.M., Adachi, K., Baumgart, J.T., Meyer, N.L., Xie, Q., Nakai, H., 2017. Adeno-associated virus (AAV) assembly-activating protein is not an essential requirement for capsid assembly of AAV serotypes 4, 5, and 11. J. Virol. 91 (3) https://doi.org/10.1128/jvi.01980-16.

Fu, S., Song, S., Liu, H., Li, Y., Li, X., Gao, X., Liang, G., 2017. ZIKA virus isolated from mosquitoes: a field and laboratory investigation in China, 2016 Sci. China Life Sci. 60 (12), 1364–1371. https://doi.org/10.1007/s11427-017-9196-8.

Gould, E., Pettersson, J., Higgs, S., Charrel, R., de Lamballerie, X., 2017. Emerging arboviruses: why today? One Health 4, 1–13. https://doi.org/10.1016/j.onehlt.2017.06.001.

Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., Suchard, M.A., 2013. Improving bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724. https://doi.org/10.1093/molbev/mss265.

Giraldo-Ramirez, S., Rendon-Marin, S., Ruiz-Saenz, J., 2020. Phylogenetic, evolutionary and structural analysis of canine parvovirus (CPV-2) antigenic variants circulating in Colombia. Viruses 12 (5), 1–15. https://doi.org/10.3390/v12050500.

Goenaga, S., Goenaga, J., Boaglio, E.R., Enria, D.A., Levis, S.D.C., 2020. Superinfection exclusion studies using west nile virus and culex flavivirus strains from Argentina. Mem. Inst. Oswaldo Cruz. 115 (5), 1–5. https://doi.org/10.1590/0074-02760200012.

Griffiths, R.C., Tavaré, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344, 403–410.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT.pdf. Nucleic Acids Symp. Ser. 41, 95–98.

Ho, S.Y.W., Phillips, M.J., Drummond, A.J., Cooper, A., 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Mol. Biol. Evol. 22, 1355–1363.

Hueffer, K., Govindasamy, L., Agbandje-McKenna, M., Parrish, C.R., 2003a. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. J. Virol. 77 (18), 10099–10105. https://doi.org/10.1128/jvi.77.18.10099-10105.2003.

Hueffer, K., Parker, J.S.L., Weichert, W.S., Geisel, R.E., Sgro, J.Y., Parrish, C.R., 2003b. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J. Virol. 77 (3), 1718–1726. https://doi.org/10.1128/jvi.77.3.1718-1726.2003.

Junglen, S., Korries, M., Grasse, W., Wieseler, J., Kopp, A., Hermanns, K., León-Juárez, M., Drosten, C., Kümmerer, B.M., 2017. Host range restriction of insect-specific flaviviruses occurs at several levels of the viral life cycle. mSphere 2, 1–15. https://doi.org/10.1128/msphere.00375-16.

Kapoor, A., Simmonds, P., Slikas, E., Li, L., Bodhidatta, L., Sethabutr, O., Delwart, E., 2010. Human bocaviruses are highly diverse, dispersed, recombination prone, and prevalent in enteric infections. J. Infect. Dis. 201 (11), 1633–1643. https://doi.org/10.1086/652416.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Khamrin, P., Okitsu, S., Ushijima, H., Maneekarn, N., 2013. Complete genome sequence analysis of novel human bocavirus reveals genetic recombination between human bocavirus 2 and human bocavirus 4. Infect. Genet. Evol. 17, 132–136. https://doi.org/10.1016/j.meegid.2013.03.040.

Kittayapong, P., Baisley, K.J., O'Neill, S.L, 1999. A mosquito densovirus infecting aedes aegypti and aedes albopictus from Thailand. Am. J. Trop. Med. Hyg. 61 (4), 612–617. https://doi.org/10.4269/ajtmh.1999.61.612.

Koppelman, M.H.G.M., Rood, I.G.H., Fryer, J.F., Baylis, S.A., Cuypers, H.T.M., 2007. Parvovirus B19 genotypes 1 and 2 detection with real-time polymerase chain reaction assays. Vox Sang. 93 (3), 208–215. https://doi.org/10.1111/j.1423-0410.2007.00957.x.

Kosakovsky.Pond, S.L., Frost, S.D.W., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.

Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk-in continuous space and time. Mol. Biol. Evol. 27, 1877–1885. https://doi.org/10.1093/molbev/msq067.

Letunic, I., Bork, P., 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46 (D1), D493–D496.

Lu, G., Wu, L., Ou, J., Li, S., 2020. Equine parvovirus-hepatitis in China: characterization of its genetic diversity and evidence for natural recombination events between the Chinese and American strains. Front. Vet. Sci. 7 (March), 1–8. https://doi.org/10.3389/fvets.2020.00121.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1 (1), 1–5.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471. https://doi.org/10.1093/molbev/msn090.

Morais, P., Pinto, J., Jorge, C.P., Troco, A.D., Fortes, F., Sousa, C.A., Parreira, R., 2020. Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: genome detection and phylogenetic characterization of viral sequences. Infect. Genet. Evol. 80, 104191 https://doi.org/10.1016/j.meegid.2020.104191.January.

Morais, P., Trovão, N.S., Abecasis, A.B., Parreira, R., 2021. Genetic lineage characterization and spatiotemporal dynamics of classical insect-specific flaviviruses: outcomes and limitations. Virus Res., 198507 https://doi.org/10.1016/j.virusres.2021.198507.

Muzyczka, N., Berns, K.I., 2001. *Parvoviridae*: the viruses and their replication. Fields Virol. 4 (69), 2327–2360.

Nei, M., Gojoborit, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Ohlund, P., Lundén, H., Blomstrom, A.L., 2019. Insect-specific virus evolution and potential effects on vector competence. Virus Genes 55 (2), 127–137. https://doi.org/10.1007/s11262-018-01629-9.

Paterson, A., Robinson, E., Suchman, E., Afanasiev, B., Carlson, J., 2005. Mosquito densonucleosis viruses cause dramatically different infection phenotypes in the C6/ 36 aedes albopictus cell line. Virology 337 (2), 253–261. https://doi.org/10.1016/j.virol.2005.04.037.

Pénzes, J.J., Soderlund-Venermo, M., Canuti, M., Eis-Hübinger, A.M., Hughes, J., Cotmore, S.F., Harrach, B., 2020. Reorganizing the family *Parvoviridae*: a revised taxonomy independent of the canonical approach based on host association. Arch. Virol. 165 (9), 2133–2146. https://doi.org/10.1007/s00705-020-04632-4.

Rambaut, A., Lam, T.T., Carvalho, L.M., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using tempest (formerly path-O-gen). Virus Evol. 2, 1–7.

Ren, X., Hoiczyk, E., Rasgon, J.L., 2008. Viral paratransgenesis in the malaria vector anopheles gambiae. PLoS Pathog. 4 (8), 4–11. https://doi.org/10.1371/journal.ppat.1000135.

Sadeghi, M., Altan, E., Deng, X., Barker, C.M., Fang, Y., Coffey, L.L., Delwart, E., 2018. Virome of >12 thousand culex mosquitoes from throughout California. Virology 523 (April), 74–88. https://doi.org/10.1016/j.virol.2018.07.029.

Sánchez-Martínez, C., Grueso, E., Carroll, M., Rommelaere, J., Almendral, J.M., 2012. Essential role of the unordered VP2 n-terminal domain of the parvovirus MVM capsid in nuclear assembly and endosomal enlargement of the virion fivefold channel for cell entry. Virology 432 (1), 45–56. https://doi.org/10.1016/j.virol.2012.05.025.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.

Shackelton, L.A., Parrish, C.R., Truyen, U., Holmes, E.C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Nat. Acad. Sci. U.S.A. 102 (2), 379–384. https://doi.org/10.1073/pnas.0406765102.

Shen, H., Zhang, W., Wang, H., Zhou, Y., Shao, S., 2015. Identification of recombination between muscovy duck parvovirus and goose parvovirus structural protein genes. Arch. Virol. 160 (10), 2617–2621. https://doi.org/10.1007/s00705-015-2541-9.

Silva, M., Morais, P., Maia, C., de Sousa, C.B., de Almeida, A.P.G., Parreira, R., 2019. A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal. Virus Res. 274, 197769 https://doi.org/10.1016/j.virusres.2019.197769.July.

Stamenkovic, G.G., Cirkovic, V.S., Siljic, M.M., Blagojevic, J.V., Knezevic, A.M., Joksic, I. D., Stanojevic, M.P., 2016. Substitution rate and natural selection in parvovirus B19. Sci. Rep. 6, 1–9. https://doi.org/10.1038/srep35759.October.

Strimmer, K., von.Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U. S. A. 94, 6815–6819.

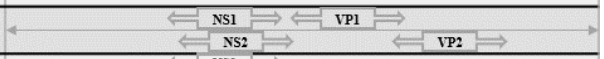
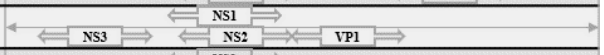
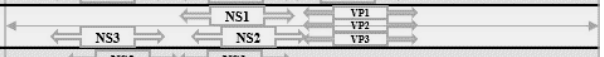
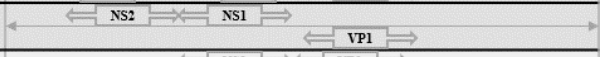
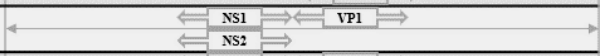
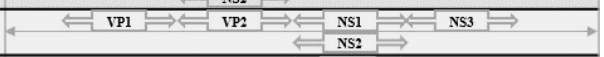
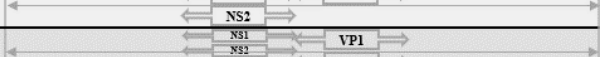
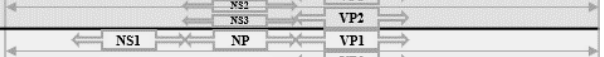
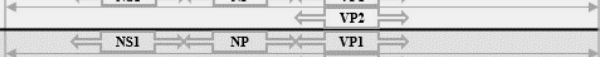
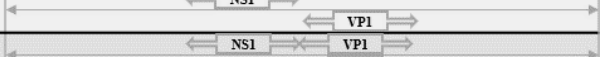
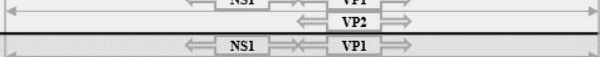
Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, 1–5.

Trifinopoulos, J., Nguyen, L., Haeseler, A. Von-Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis 44, 232–235. https://doi:10.1093/nar/gkw256.

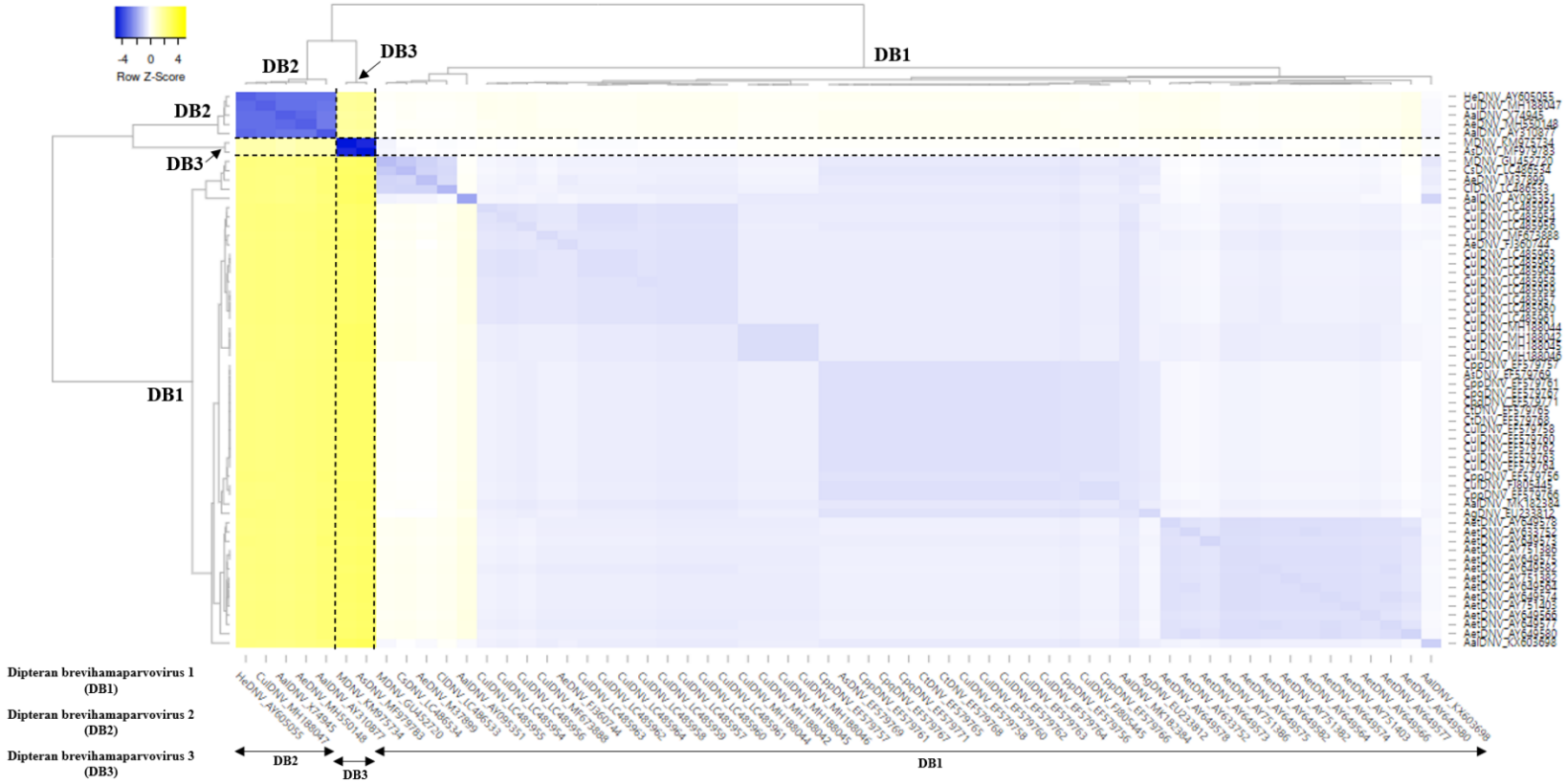
Trovao, N.S., Baele, G., Vrancken, B., Bielejec, F., Suchard, M.A., Fargette, D., Lemey, P., 2015. Host ecology determines the dispersal patterns of a plant virus. Virus Evol. 1 (1) https://doi.org/10.1093/ve/vev016.

Yang, W.T., Shi, S.H., Jiang, Y.L., Zhao, L., Chen, H.L., Huang, K.Y., Wang, C.F., 2016. Genetic characterization of a densovirus isolated from great tit (parus major) in China. Infect. Genet. Evol. 41, 107–112. https://doi.org/10.1016/j.meegid.2016.03.035

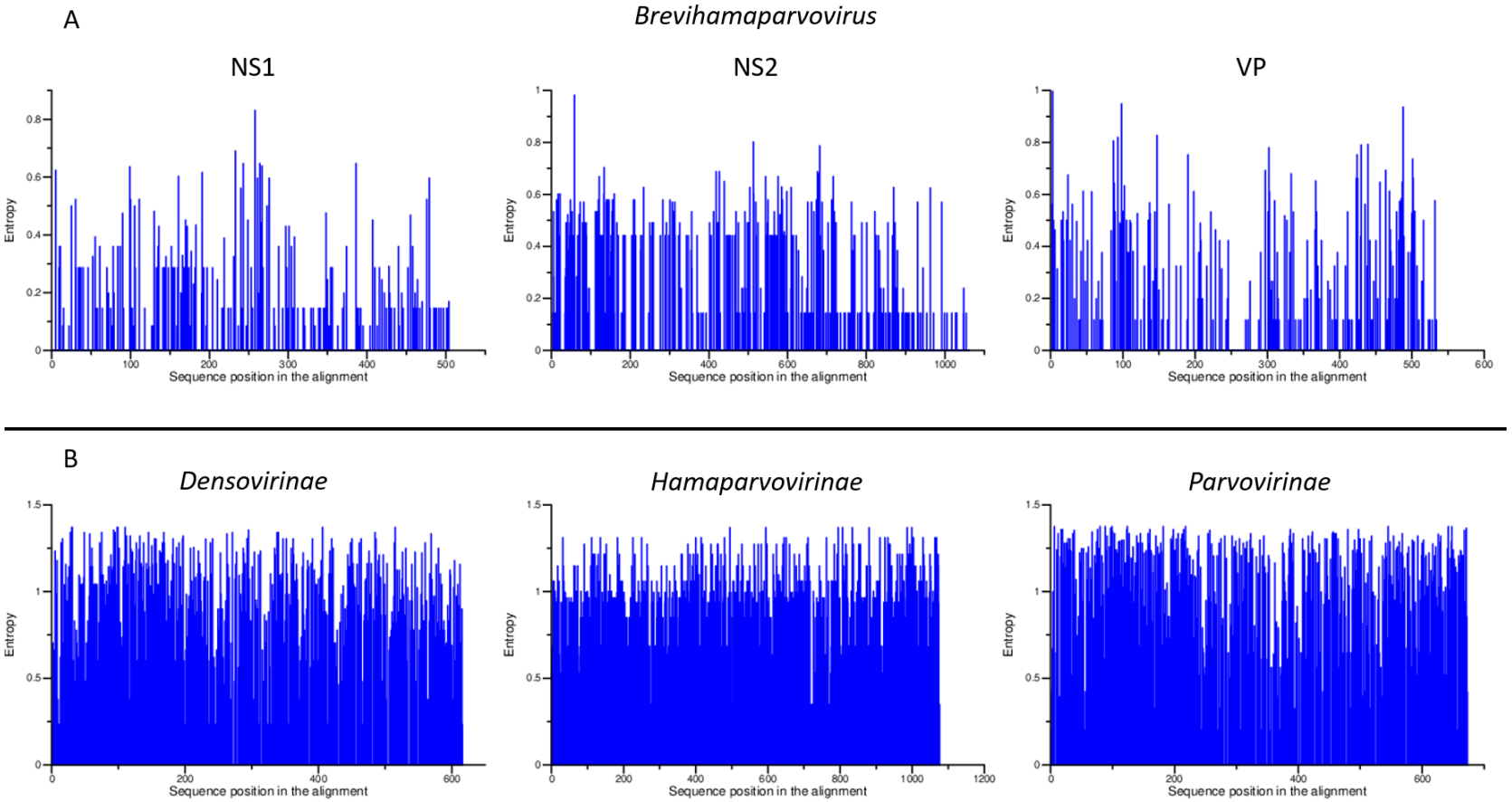
Supplementary Fig. 1: Schematic representation of nucleotide sequences from each genus in the *Parvoviridae* family with different ORFs identified. Not representative of the size of each ORF, only their organization and sequence. NS – non-structural protein; VP – viral protein; NP – nucleoprotein; AAP – assembly activating protein.

Accession number	Subfamily	Genus	Total number of base pairs	Genomic structure
NC_034532	*Densovirinae*	*Ambidensovirus*	4981	
NC_022564	*Densovirinae*	*Miniambidensovirus*	4945	
NC_026943	*Densovirinae*	*Aquambidensovirus*	6334	
NC_005041	*Densovirinae*	*Blattambidensovirus*	5335	
NC_011545	*Densovirinae*	*Hepandensovirus*	6310	
NC_018450	*Densovirinae*	*Iteradensovirus*	5053	
NC_002190	*Densovirinae*	*Penstyldensovirus*	3909	
NC_000936	*Densovirinae*	*Pefuambidensovirus*	5454	
NC_004286	*Densovirinae*	*Protoambidensovirus*	6039	
NC_022748	*Densovirinae*	*Scindoambidensovirus*	5280	
KX603698	*Hamaparvovirinae*	*Brevihamaparvovirus*	4073	
NC_040843	*Hamaparvovirinae*	*Chaphamaparvovirus*	4442	
MN049932	*Hamaparvovirinae*	*Ichthamaparvovirus*	4001	
NC_039043	*Hamaparvovirinae*	*Penstylhamaparvovirus*	3873	
NC_034445	*Parvovirinae*	*Amdoparvovirus*	4242	
NC_040603	*Parvovirinae*	*Aveparvovirus*	5515	
NC_038535	*Parvovirinae*	*Bocaparvovirus*	5263	
NC_006259	*Parvovirinae*	*Copiparvovirus*	5610	
NC_027429	*Parvovirinae*	*Dependoparvovirus*	4590	
NC_038540	*Parvovirinae*	*Erythroparvovirus*	4986	
NC_038545	*Parvovirinae*	*Protoparvovirus*	4936	
NC_038546	*Parvovirinae*	*Tetraparvovirus*	5114	

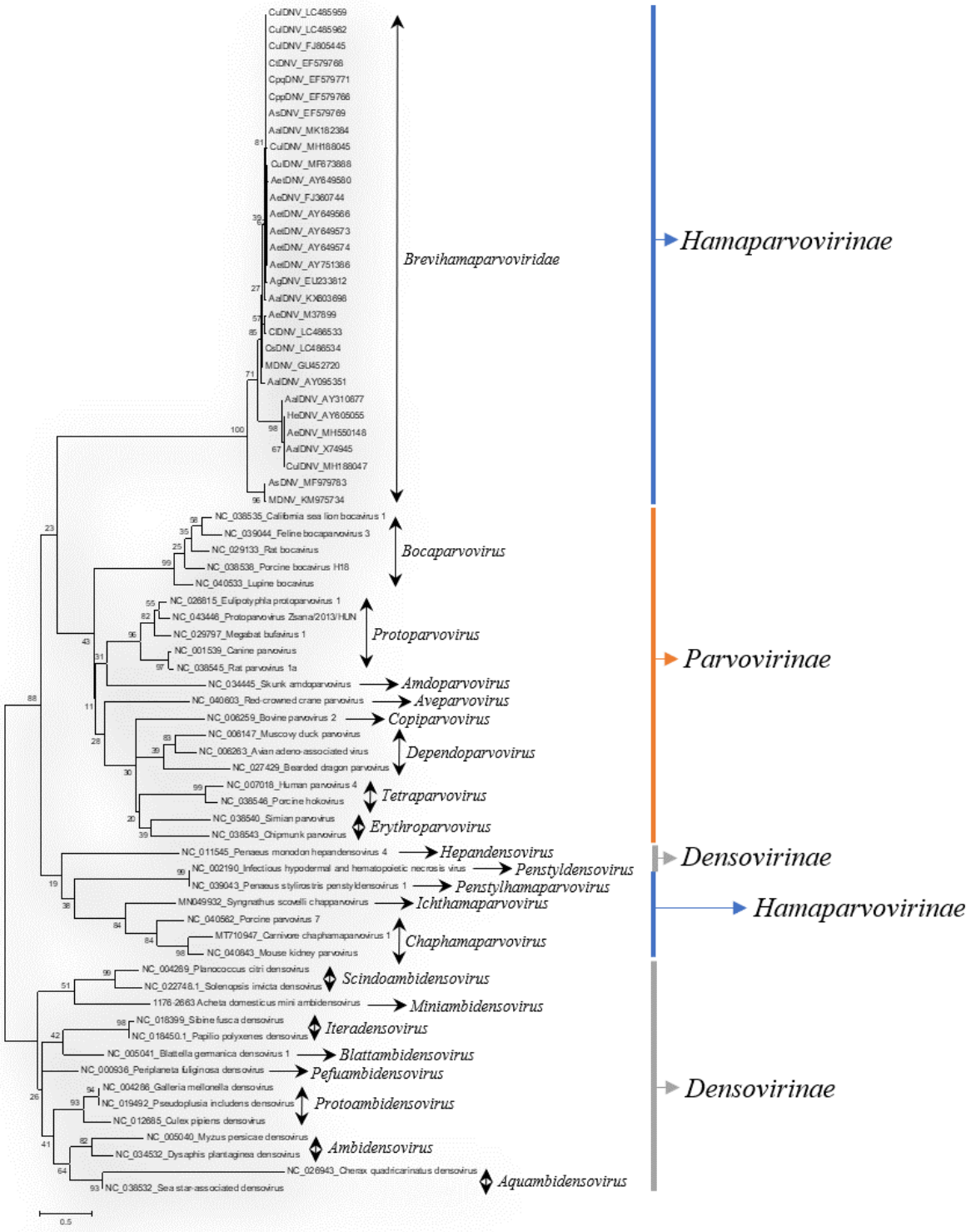
Supplementary Fig. 2: Heat map representing inter-sequence genetic diversity of *Brevihamaparvovirus*. Representative tree obtained on IQ-TREE (maximum likelihood, GTR+Γ+I model) based on NS1 nt sequences (reported in Supplementary Table 1), and Z-Scores estimated based on pairwise evolutionary distances using MegaX.



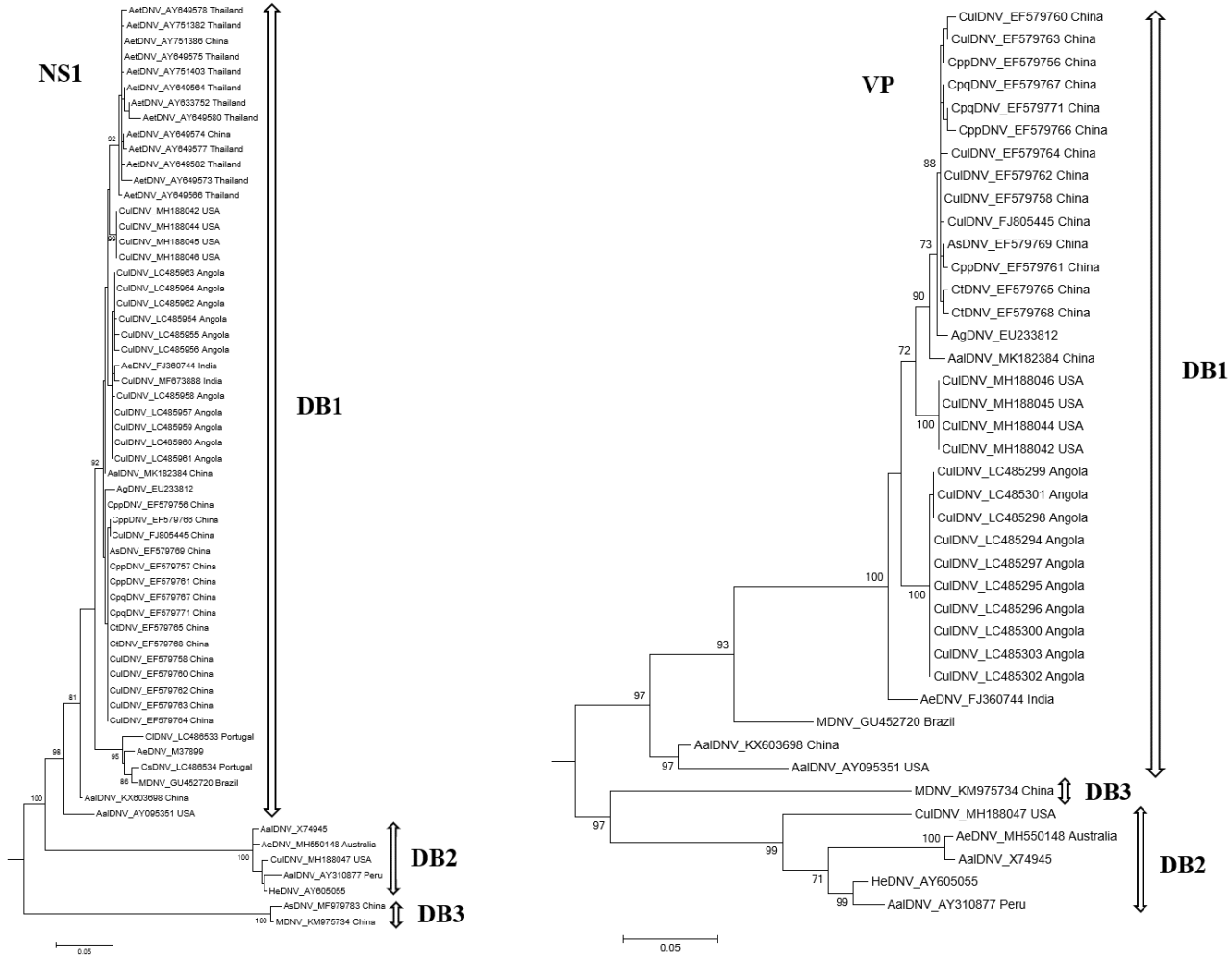
Supplementary Fig. 3: (A) Entropy on the basis of the Shannon function (Shannon entropy-one) for different ORF-coding sequences of *Brevihamaparvovirus*; (B) Entropy on the basis of Shannon function (Shannon entropy-one) for NS1-coding sequences of different subfamilies in the *Parvoviridae* family.



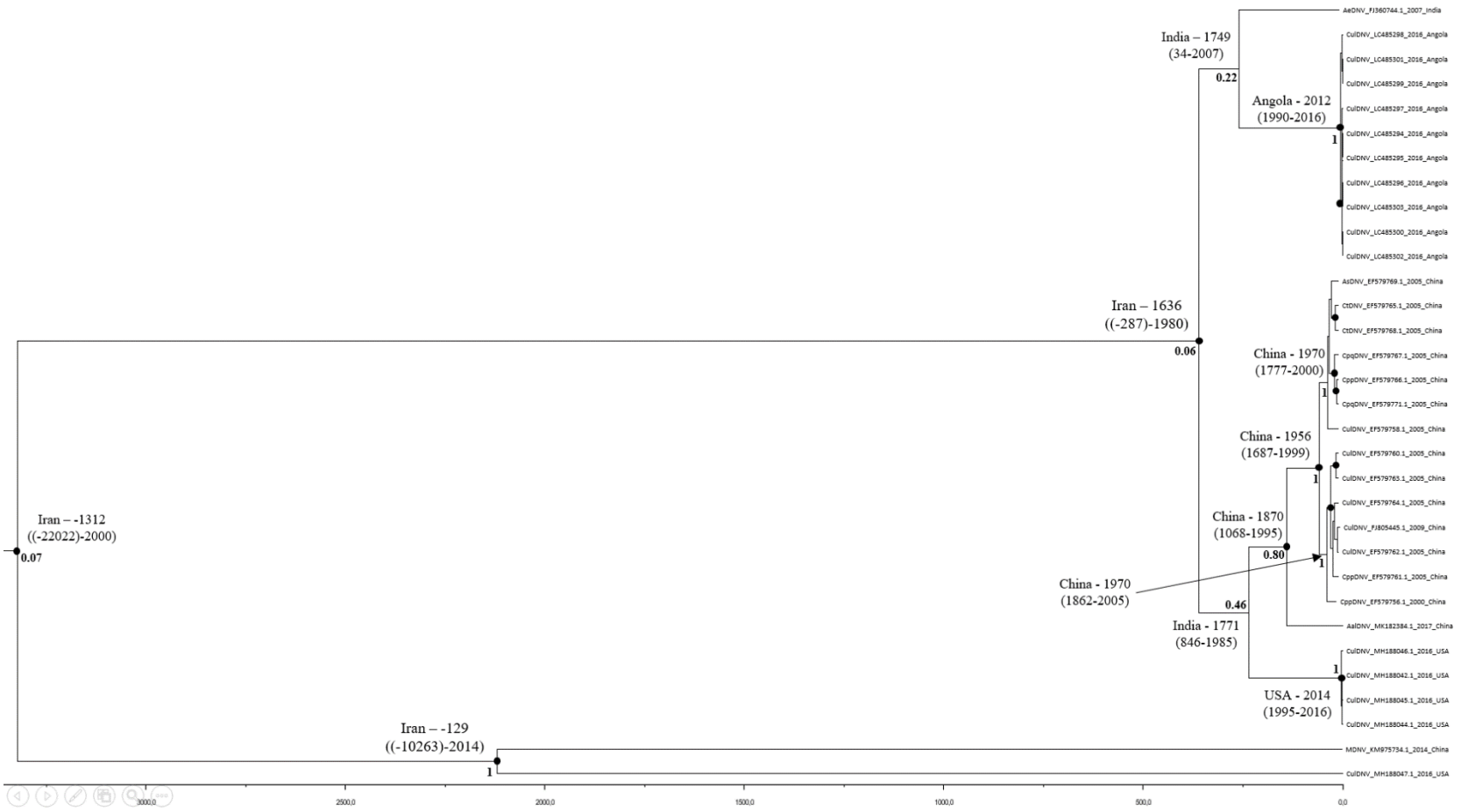
Supplementary Fig. 4: NS1 maximum likelihood phylogenetic tree of several parvovirus genera and subfamilies, estimated under a WAG substitution model using IQ-TREE (phylogeny test with 1000 bootstrap replications). Isolates are shown in Supplementary Table 2.



Supplementary Fig. 5: Maximum likelihood tree of *Brevihamaparvovirus* NS1 and VP nucleotide sequences, estimated under a GTR+Γ+I substitution model using IQ-TREE (phylogeny test with 1000 bootstrap replications). The different genetic lineages (DB1-3) are indicated.



Supplementary Fig. 6: Continuous phylogeographic analysis of *Brevihamaparvovirus* VP coding sequence. At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles, while the decimals associated with certain nodes indicate the inferred location PP.



Supplementary Table 1: Brevihamaparvovirus sequences used in present study. Included ORFs identified for each sequence (NS1 – non-structural protein 1; NS2 – non-structural protein 2; VP – viral protein). The taxonomy designations follow the ICTV classification as of July 2021.

Accession number	Species	Nomenclature	Abbreviation	ORFs	Host species	Geographical origin	Collection Year
AY095351[a x]	unclassified Brevihamaparvovirus	Aedes albopictus C6/36 cell densovirus	AalDNV	NS1, NS2, VP	*Aedes albopictus*	Unknown	Unknown
AY310877[a x]	unclassified Brevihamaparvovirus	Aedes albopictus C6/36 cell densovirus	AalDNV	NS1, NS2, VP	*Aedes albopictus*	Unknown	Unknown
AY605055[a x]	unclassified Brevihamaparvovirus	Haemagogus equinus densovirus	HeDNV	NS1, NS2, VP	*Haemagogus equinus*	Unknown	Unknown
AY633752	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649564	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649566	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649573	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649574	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649577	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993

AY649578	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649579	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649580	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY649582	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY751382	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY751386	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
AY751403	unclassified Brevihamaparvovirus	Aedes aegypti Thai densovirus	AetDNV	NS1	*Aedes aegypti Thai*	Thailand	1993
EF579756	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CppDNV	NS1, NS2, VP	*Culex pipiens pallens*	China	2000
EF579757	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CppDNV	NS1, NS2	*Culex pipiens pallens*	China	2005
EF579758	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	China	2005
EF579760	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	China	2005
EF579761	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CppDNV	NS1, NS2, VP	*Culex pipiens pallens*	China	2005

EF579762	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	China	2005
EF579763	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	China	2005
EF579764	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	China	2005
EF579765	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CtDNV	NS1, NS2, VP	*Culex tritaeniorhynchus*	China	2005
EF579766	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CppDNV	NS1, NS2, VP	*Culex pipiens pallens*	China	2005
EF579767	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CpqDNV	NS1, NS2, VP	*Culex pipiens quinquefasciatus*	China	2005
EF579768	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CppDNV	NS1, NS2, VP	*Culex tritaeniorhynchus*	China	2005
EF579769	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	AsDNV	NS1, NS2, VP	*Anopheles sinensis*	China	2005
EF579771	Dipteran brevihamaparvovirus 1	Culex pipiens pallens densovirus	CpqDNV	NS1, NS2, VP	*Culex pipiens quinquefasciatus*	China	2005

EU233812[b]	Dipteran brevihamaparvovirus 1	Anopheles gambiae densonucleosis virus	AgDNV	NS1, NS2, VP	*Anopheles gambiae*	Unknown	Unknown
FJ360744	Dipteran brevihamaparvovirus 1	Aedes aegypti densovirus	AeDNV	NS1, NS2, VP	*Aedes aegypti*	India	2007
FJ805445	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	China	2009
GU452720[a]	unclassified Brevihamaparvovirus	Mosquito densovirus	MDNV	NS1, NS2, VP	Unknown	Unknown	Unknown
KM975734	unclassified Brevihamaparvovirus	Mosquito densovirus	MDNV	NS1, NS2, VP	Unknown	China	2014
KX603698	unclassified Brevihamaparvovirus	Aedes albopictus densovirus	AalDNV	NS1, NS2, VP	*Aedes albopictus*	China	2015
LC485954 (NS1)/LC485294 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485955 (NS1)/LC485295 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485956 (NS1)/LC485296	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485957 (NS1)/LC485297 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016

LC485958 (NS1)/LC485298 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485959 (NS1)/LC485299 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485960 (NS1)/LC485300 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485961 (NS1)/LC485301 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485962 (NS1)/LC485302 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485963 (NS1)/LC485303 (VP)	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, VP	*Culex sp.*	Angola	2016
LC485964	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1	*Culex sp.*	Angola	2016
LC486533	unclassified Brevihamaparvovirus	Culex laticinctus densovirus	ClDNV	NS1	*Culex laticinctus*	Portugal	2018

LC486534	unclassified Brevihamaparvovirus	Culiseta longlareolata densovirus	CsDNV	NS1	*Culiseta longlareolata*	Portugal	2018
M37899	Dipteran brevihamaparvovirus 1	Aedes aegypti densovirus	AeDNV	NS1	Unknown	Unknown	Unknown
MF673888	unclassified Brevihamaparvovirus	Mosquito densovirus	CulDNV	NS1	*Culex sp.*	India	2015
MF979783	unclassified Brevihamaparvovirus	Mosquito densovirus	AsDNV	NS1	*Armigeres subalbatus*	China	2016
MH188042	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	USA	2016
MH188044	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	USA	2016
MH188045	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	USA	2016
MH188046	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	USA	2016
MH188047	unclassified Brevihamaparvovirus	Culex densovirus	CulDNV	NS1, NS2, VP	*Culex sp.*	USA	2016
MH550148[c]	Dipteran brevihamaparvovirus 1	Aedes albopictus densovirus	AeDNV	NS1, NS2, VP	Aedes albopictus	Unknown	Unknown
MK182384	unclassified Brevihamaparvovirus	Aedes albopictus densovirus	AalDNV	NS1, NS2, VP	Aedes albopictus	China	2017

X74945[a]	Dipteran brevihamaparvovirus 2	Aedes albopictus densovirus 2	AalDNV	NS1, NS2, VP	Aedes albopictus	Unknown	Unknown

; [a]Sequences obtained via C6/36 (Aedes albopictus) cell culture; [b]Sequences obtained via Sua5B (Anopheles gambiae) cell culture; [c]Sequences obtained via Aag2 (*Aedes aegypti*) cell culture.

Supplementary Table 2: Information regarding all virus sequences from the *Parvoviridae* family employed in present study (taxonomy based on the ICTV classification scheme as of July 2021).

Accession number	**Subfamily**	**Genus**	**Species**	**Nomenclature**	**Host species**	**Geographical origin**	**Collection Year**
NC_034532	*Densovirinae*	*Ambidensovirus*	Hemipteran ambidensovirus 2	Dysaphis plantaginea densovirus	*Dysaphis plantaginea*	United Kingdom	2002
NC_019492	*Densovirinae*	*Ambidensovirus*	Lepidopteran ambidensovirus 1	Pseudoplusia includens densovirus	*Pseudoplusia includens*	Canada	1985
NC_005040	*Densovirinae*	*Ambidensovirus*	Hemipteran ambidensovirus 3	Myzus persicae densovirus	*Myzus persicae*	Unknown	Unknown
NC_022564	*Densovirinae*	*Miniambidensovirus*	Orthopteran miniambidensovirus 1	Acheta domesticus mini ambidensovirus	*Acheta domesticus*	USA	2012
NC_038532	*Densovirinae*	*Aquambidensovirus*	Asteroid aquambidensovirus 1	Sea star-associated densovirus	*Asterias amurensis*	USA	2013
NC_026943	*Densovirinae*	*Aquambidensovirus*	Decapod aquambidensovirus 1	Cherax quadricarinatus densovirus	*Cherax quadricarinatus*	Australia	1999

NC_005041	*Densovirinae*	*Blattambidensovirus*	Blattodean blattambidensovirus 1	Blattella germanica densovirus 1	*Blattella germanica*	USA	Unknown
NC_011545	*Densovirinae*	*Hepandensovirus*	Decapod hepandensovirus 1	Penaeus monodon hepandensovirus 4	*Penaeus monodon*	India	Unknown
NC_018450	*Densovirinae*	*Iteradensovirus*	Lepidopteran iteradensovirus 4	Papilio polyxenes densovirus	*Papilio polyxenes*	Canada	2011
NC_018399	*Densovirinae*	*Iteradensovirus*	Lepidopteran iteradensovirus 2	Sibine fusca densovirus	*Sibine fusca*	Colombia	1977
NC_002190	*Densovirinae*	*Penstyldensovirus*	Decapod penstyldensovirus 1	Infectious hypodermal and hematopoietic necrosis virus	Penaeid shrimp	Unknown	Unknown
NC_000936	*Densovirinae*	*Pefuambidensovirus*	Blattodean pefuambidensovirus 1	Periplaneta fuliginosa densovirus	*Periplaneta fuliginosa*	Unknown	Unknown
NC_004286	*Densovirinae*	*Protoambidensovirus*	Lepidopteran protoambidensovirus 1	Galleria mellonella densovirus	*Galleria mellonella*	Unknown	Unknown
NC_012685	*Densovirinae*	*Protoambidensovirus*	Dipteran protoambidensovirus 1	Culex pipiens densovirus	*Culex pipiens*	France	1998
NC_022748	*Densovirinae*	*Scindoambidensovirus*	Hymenopteran scindoambidensovirus 1	Solenopsis invicta densovirus	*Solenopsis invicta*	Argentina	2008
NC_004289	*Densovirinae*	*Scindoambidensovirus*	Hemipteran scindoambidensovirus 1	Planococcus citri densovirus	*Planococcus citri*	Unknown	Unknown

MT710947	*Hamaparvovirinae*	*Chaphamaparvovirus*	Carnivore chaphamaparvovirus 1	Carnivore chaphamaparvovirus 1	Dog	Italy	2019
NC_040843	*Hamaparvovirinae*	*Chaphamaparvovirus*	Rodent chaphamaparvovirus 1	Mouse kidney parvovirus	*Mus musculus*	Australia	2016
NC_040562	*Hamaparvovirinae*	*Chaphamaparvovirus*	Ungulate chaphamaparvovirus 1	Porcine parvovirus 7	*Sus scrofa domesticus*	China	2015
MN049932	*Hamaparvovirinae*	*Ichthamaparvovirus*	Syngnathid ichthamaparvovirus 1	Syngnathus scovelli chapparvovirus	*Syngnathus scovelli*	USA	2016
NC_039043	*Hamaparvovirinae*	*Penstylhamaparvovirus*	Decapod penstylhamaparvovirus 1	Penaeus stylirostris penstyldensovirus 1	*Litopenaeus stylirostris*	USA	1998
NC_034445	*Parvovirinae*	*Amdoparvovirus*	Carnivore amdoparvovirus 4	Skunk amdoparvovirus	*Mephitis mephitis*	Canada	2014
NC_040603	*Parvovirinae*	*Aveparvovirus*	Gruiform aveparvovirus 1	Red-crowned crane parvovirus	*Grus japonensis*	China	2014
NC_038535	*Parvovirinae*	*Bocaparvovirus*	Pinniped bocaparvovirus 1	California sea lion bocavirus 1	*Zalophus californianus*	USA	2010
NC_038538	*Parvovirinae*	*Bocaparvovirus*	Ungulate bocaparvovirus 4	Porcine bocavirus H18	*Sus scrofa*	China	2008
NC_039044	*Parvovirinae*	*Bocaparvovirus*	Carnivore bocaparvovirus 5	Feline bocaparvovirus 3	*Felis catus*	USA	2013

NC_029133	*Parvovirinae*	*Bocaparvovirus*	Rodent bocaparvovirus 1	Rat bocavirus	*Rattus norvegicus*	Hong Kong	2010
NC_040533	*Parvovirinae*	*Bocaparvovirus*	unclassified Bocaparvovirus	Lupine bocavirus	*Canis lupus signatus*	Portugal	2011
NC_006259	*Parvovirinae*	*Copiparvovirus*	Ungulate copiparvovirus 1	Bovine parvovirus 2	Bovine	Unknown	Unknown
NC_027429	*Parvovirinae*	*Dependoparvovirus*	Squamate dependoparvovirus 2	Bearded dragon parvovirus	*Pogona vitticeps*	Hungary	2009
NC_006147	*Parvovirinae*	*Dependoparvovirus*	Anseriform dependoparvovirus 1	Muscovy duck parvovirus	*Cairina moschata*	Unknown	Unknown
NC_006263	*Parvovirinae*	*Dependoparvovirus*	Avian dependoparvovirus 1	Avian adeno-associated virus	Avian	Unknown	Unknown
NC_038540	*Parvovirinae*	*Erythroparvovirus*	Primate erythroparvovirus 2	Simian parvovirus	*Macaca fascicularis*	Unknown	Unknown
NC_038543	*Parvovirinae*	*Erythroparvovirus*	Rodent erythroparvovirus 1	Chipmunk parvovirus	*Tamias sibiricus asiaticus*	South Korea	Unknown
NC_038545	*Parvovirinae*	*Protoparvovirus*	Rodent protoparvovirus 2	Rat parvovirus 1a	Rat	Unknown	Unknown
NC_029797	*Parvovirinae*	*Protoparvovirus*	Chiropteran protoparvovirus 1	Megabat bufavirus 1	*Pteropus vampyrus*	Indonesia	2012
NC_026815	*Parvovirinae*	*Protoparvovirus*	Eulipotyphla protoparvovirus 1	Eulipotyphla protoparvovirus 1	*Crocidura hirta*	Zambia	2012

NC_001539	*Parvovirinae*	*Protoparvovirus*	Carnivore protoparvovirus 1	Canine parvovirus	Dog	Unknown	Unknown
NC_043446	*Parvovirinae*	*Protoparvovirus*	Ungulate protoparvovirus 2	Protoparvovirus Zsana/2013/HUN	*Sus scrofa domesticus*	Hungary	2013
NC_038546	*Parvovirinae*	*Tetraparvovirus*	Ungulate tetraparvovirus 2	Porcine hokovirus	Porcine	China	Unknown
NC_007018	*Parvovirinae*	*Tetraparvovirus*	Primate tetraparvovirus 1	Human parvovirus 4	*Homo sapiens*	USA	Unknown

Supplementary Table 3: Assessment of selective pressure of brevihamaparvovirus using three different methods (Single Likelihood Ancestor Counting or SLAC and Fixed Effects Likelihood or FEL available in the DataMonkey server, and by the Synonymous Non-synonymous Analysis Program or SNAP, hosted the HIV Los Alamos Database) for each coding region, with a p-value of 0.05. Percentages shown for number of sites, either for negative or positive selection, to the total number of sites for each genomic region.

GO	NS1	NS2	VP
FEL	0.82	0.53	0.11
SLAC	0.84	0.57	0.13
SNAP	0.64	0.68	0.14
Sites under negative selection	**NS1**	**NS2**	**VP**
Number of sites	**168**	**354**	**178**
FEL	15 (9%)	40 (11%)	67 (38%)
SLAC	4 (2%)	5 (1%)	26 (15%)
Sites under positive selection	**NS1**	**NS2**	**VP**
Number of sites	**168**	**354**	**178**
FEL	0	2 (0.6%)	0
SLAC	0	0	0

Supplementary Table 4: (A) Comparative analysis of the performance of parametric and non-parametric coalescent priors in the analysis of brevihamaparvoviruses VP-coding sequences. Best run highlighted as bold, as well as with Bayes factor values for all other runs; (B) Evaluation of rates for coalescent combined with different geographic diffusion priors for the VP-coding region of brevihamaparvoviruses. Best run results are highlighted as bold, as well as with Bayes factor values for all other runs (in brackets).

	root_age [95% HPD]	**mean rate [95% HPD]**	**stdev [95% HPD]**	**coefficient of variation [95% HPD]**	**PS1**	**PS2**	**SS1**	**SS2**
A								
Constant	-5.79E21 [-2.7E11-2000]	1.16E-3 [2.60E-35-6.88E-3]	3.87E-2 [4.9E-324-0.1511]	0.54 [0-3.2]	-1776 (13)	-1778 (15)	-1776 (13)	-1778 (15)
Exponential	1345 [359-2000]	2.24E-4 [2.16E-5-4.51E-4]	4.04E-6 [2.16E-5-4.52E-4]	4.00E-3 [0-1.65E-15]	-1778 (15)	-1778 (15)	-1778 (15)	-1778 (15)
Expansion	-162 [-423-1821]	2.24E-4 [2.12E-05-4.51E-4]	4.01E-6 [4.9E-324-1.76E-15]	4E-3 [0-1.65E-15]	-1780 (17)	-1784 (21)	-1781 (18)	-1785 (22)
GMRF Skyride	1992 [1978-2000]	3.72E-3 [1.15E-3-6.43E-3]	0.18 [7.77E-4-0.69]	3.01 [2.03-4.23]	-1776 (13)	-1776 (13)	-1776 (13)	-1776 (13)
Skygrid	-1.1E61 [-55165-2000]	1.12E-3 [1.02E-62-6.84E-3]	3.81E-2 [4.9E-324-0.15]	0.49 [0-3.17]	-1767 (4)	-1770 (7)	-1768 (5)	-1771 (6)
Skyline	-4.07E6 [-1600-2000]	1.30E-3 [5.17E-13-8.64E-3]	1.12E-2 [4.9E-324-0.05]	0.32 [0-2.33]	-1769 (6)	**-1763**	-1769 (6)	**-1763**
B								
Skyline+Brownian	1999 [1999-2000]	1,01E-02 [7,03E-03-0,01]	0,53 [0,02-1,60]	2,38 [1,40-3,32]	-1975 (195)	-1975 (195)	-1968 (188)	-1967 (187)
Skyline+CauchyRRW	-1486 [-10014-2000]	3,33E-03 [1,08E-07-0,01]	9,83E-02 [4,9E-324-0,51]	0,60 [0-2,50]	-1805 (25)	-1808 (28)	-1809 (29)	-1808 (28)
Skyline+LognormalRRW	nd	nd	nd	nd	nd	nd	nd	nd
Skyline+GammaRRW	-3150 [-23724-2000]	5,64E-04 [3,19E-88-3,22E-04]	1,42E-02 [4,9E-324-1,00E-03]	0,196 [0-1,89]	**-1780**	-1784 (4)	-1797 (17)	-1800 (20)

Supplementary Table 5: Pairwise genetic distances between all brevihamaparvoviruses NS1 aa sequences. Excel spreadsheet available online at: https://doi:10.1016/j.virusres.2022.198728

Chapter 7. Supplementary Results

Author summary

Through the course of this project, extensive data was obtained following the analysis of viral sequences representing three viral genera, from as many viral families [*Flavivirus* (*Flaviviridae*), *Alphamesonivirus* (*Mesoniviridae*), and *Brevihamaparvovirus* (*Parvoviridae*)], which we identified in Chapter 1 as some of the most relevant today, as far as ISV research is concerned. While the most important data obtained in the course of this study has already been presented in the previous chapters as research papers, supplementary data, also deemed important in the context of our analysis, will be incorporated in this chapter. This will include (i) the results of mesonivirus-targeted viral surveys in mosquito pools collected in Portugal and Angola, (ii) a more detailed view of selective pressure and temporal signal analyses and (iii) an attempt to evaluate all cISF spatiotemporal dispersal (not possible in past research, as conveyed in chapter 4) using other flaviviruses sequences.

1. Mesonivirus viral surveys

1.1. Introduction

Parallel to the detection of insect-specific flaviviruses and brevihamaparvoviruses in mosquito pools collected in both Portugal (50 pools; Chapter 2) and Angola (20 pools; Chapter 3), we also sought to detect mesoniviruses sequences in these same mosquito pools. Had this endeavor been successful, it would have resulted in the first characterization of mesoniviruses sequences from both these countries. However, no primers that would enable such detection were available at the beginning of this thesis project in IHMT nor available in literature with the coverage we wanted, and for that reason they had to be designed. In addition, assessment of their performance required the development of a positive control.

1.2. Material and Methods

Mesonivirus sequence detection was performed on all 50 mosquito pools from mosquitoes collected in Portugal (see Supplementary Table 2 from Chapter 2) and on 20

of a total of 60 mosquito pools from mosquitoes collected in Angola. At this stage, multiple alignments of nt sequences from all available mosquito mesoniviruses sequences (from Supplementary Table 1 in Chapter 5) were performed using the iterative G-INS-i method as implemented in MAFFT vs. 7 (Katoh and Standley, 2013). Primer design was carried out using Primer Design M (Yoon & Leitner, 2015) where conserved regions, with low sequence entropy and the highest possible Tm, were selected as their potential hybridization targets, considering a region-of-interest (amplicon size) of 1800 base pairs. We selected two pairs of primers that would allow the amplification (via nested PCR) of a partial sequence inside the RdRp coding region. Primer sequences as well as the thermal profiles used for PCR (which were defined *de novo* and improved alongside experimental work) can be viewed in Table 1.

Table 1: Primer sequences, and thermal profiles used for partial sequence amplification of the mesonivirus RdRp-coding region; indicated nt coordinates correspond to those in the Houston virus sequence (accession number MH719099).

Target gene	Primer sequence (5'-3')	Coordinates (MH719099)	Thermocycling conditions
RdRp (Mesoniviruses)	1st PCR		
	MesonV_F1: GATTATCCHAAATGGGAYCGYCG	10.079-10.101 (+)	95 ºC - 2 min; 25 cycles [95 ºC – 30 seg; 50 ºC - 2 min; 72 ºC – 2 mins]; 25 cycles [95 ºC – 30 seg; 50 ºC - 2 min; 72 ºC – 2 min+5 seg per cycle]; 72ºC – 10 min
	MesonV_R1: GGRATTTGKGTGTCAGTTWGCCATATATG	11.952-11.924 (-)	
	2nd PCR		
	MesonV_F2: CCTGARTTTGGACGCATGTAYWCC	10.733-10.756 (+)	95 ºC - 2 min; 50 cycles [95 ºC – 30 seg; 50 ºC - 2 min; 72 ºC – 1 min]; 72ºC – 10 min
	MesonV_R2: GCATAATTARTTGRTGATATGGTCTGAC	11.717-11.690 (-)	

PCR amplifications were carried out in final volumes of 20 µL (10 µL of PCR Mix), using a total of 30pmol of each degenerate primer mixture (forward and reverse) and 4 µL of cDNA

Detection of mesonivirus sequences was attempted using cDNA obtained as previously described (Carapeta et al., 2015; Pimentel et al., 2019) from mosquito pools used in Chapter 2 and 3 of this thesis. In the absence of an adequate positive control (viral isolate), primer detection performance was tested using an artificial molecule, also designed in the context of this work. It corresponded to a double stranded DNA sequence with the expected amplicon size of 1800 nt including a heterologous sequence (*Bacillus subtilis*

bacteriophage SPP1 DNA, used as stuffer DNA), flanked by viral nt sequences of the Houston virus genome (isolate HOUV-M742, accession number MH719099).

New primers were created in order to obtain the positive control fragment (Table 2) in the form of a PCR amplicon when phage SPP1 genomic DNA was used as the amplification template. These new set of primers allowed for the hybridization of 20 nt at their 3' half with DNA from the bacteriophage SPP1 (accession number X97918), while their 5' half displayed target sequences for the Houston virus (as a representative of the viruses we wished to detect). After PCR, aliquots of the amplification reactions were analyzed by agarose gel electrophoresis (1%) which allowed the detection of the expected 1.8 kb DNA fragment. The Zymoclean Gel DNA Recovery Kit (Zyma Research) was used to purify the DNA fragment from the agarose gel. Serial dilutions of the purified DNA fragment were used, along with the newly designed primers, to template the amplification of the intended target DNA. Primer sequences and the thermal profiles used are indicated in Table 1.

Table 2: Thermal profiles used for PCR and sequences of new designed primers specific to both nucleotide SPP1 sequence and Houston virus isolate; sequence that allows for hybridization with DNA from the bacteriophage SPP1 in bold.

Target sequences	**Primer sequence (5'-3')**	**Thermocycling conditions**
SPP1 (X97918) + Houston virus isolate (MH719099)	SPPMeson_F1: GATTATCCAAAATGGGATCGCCG**CCTCCTTGATCTCGTGAGACG** SPPMeson_R1: **GGAATTTGTGTGTCAGTTTGC**CATATATGGACCCTGTCTTTCTGCTCTCC	95 ºC - 3 min; 45 cycles [95 ºC – 30 seg; 55 ºC – 30 seg; 72 ºC – 2 mins]; 72ºC – 5 min

1.3. Results and Discussion

We were unable to detect any mesonivirus sequences from mosquito pools from both Portugal and Angola. We were, however, able to demonstrate that our newly designed primers could potentially detect mesonivirus genome sequences, as demonstrated by amplification of the intended amplification product (for the first round) using as template 1:1000 and 1:10000 serial dilutions of the purified positive control amplicon. Even though this positive control was only able to assess the ability of the newly designed

MesonV_F1 and MesonV_R1 primer to detect a specific viral sequence (Houston virus), the degeneracy of those same primers should allow for the detection of a higher number of viral sequences, if present. However, sequence detection is more sensitive when the intended target is small, and decreasing as the amplicons grow bigger (Habbal et al., 2009). Therefore, the detection of a relatively large amplicon (1.8 kb) may have been compromised not only by the presence/absence of mesonivirus genomes, but also by the quality of the viral DNA. Its putative fragmentation or degradation, could have also hindered mesonivirus sequence detection.

2. Selective pressure analyses

2.1. Introduction

Selective pressure forces have been key in evolution of viral strains over time (Ghosh & Chakraborty, 2020). Indeed, these forces could be key in how viral genomes adapt over time when pressured by changing environments. Even though the genomes of the three ISV groups we investigated (cISF, Chapter 4; mesonivirus, Chapter 5; brevihamaparvovirus, Chapter 6) look to evolve under strong purifying pressure, mostly indicating that evolutionary drive is against amino acid change, we did identify a few numbers of sites in the viral genome that seem to be under positive selection. In this section we will look into these sites in a more detailed manner, checking how amino acids in these sites diverge between different ISV sequences and comparing the locations where the codons encoding the changing amino acids (diversifying selection) occur.

2.2. Material and Methods

These analyses were performed simultaneously, and with the same tools, as those mentioned in section 2.5 of Chapter 4. For this specific case, we considered the results of the analyses of complete genome sequence alignments for all three ISV groups (cISF, mesonivirus and brevihamaparvovirus) in the DataMonkey webserver. We identified the location of the sites under strong positive selection ($\omega > 1$; $p < 0,05$) in both FEL and SLAC analyses using coordinates available in the GenBank database records.

2.3. Results and Discussion

For cISF, while most of the genome seems to be evolving under strong purifying selection (more detailed results in Supplementary Table 1 in Chapter 4), high ω values were observed especially in the NS2a coding region (Fig. 1), but also in a minor number of selected positions elsewhere (e.g. NS2b or NS4a).

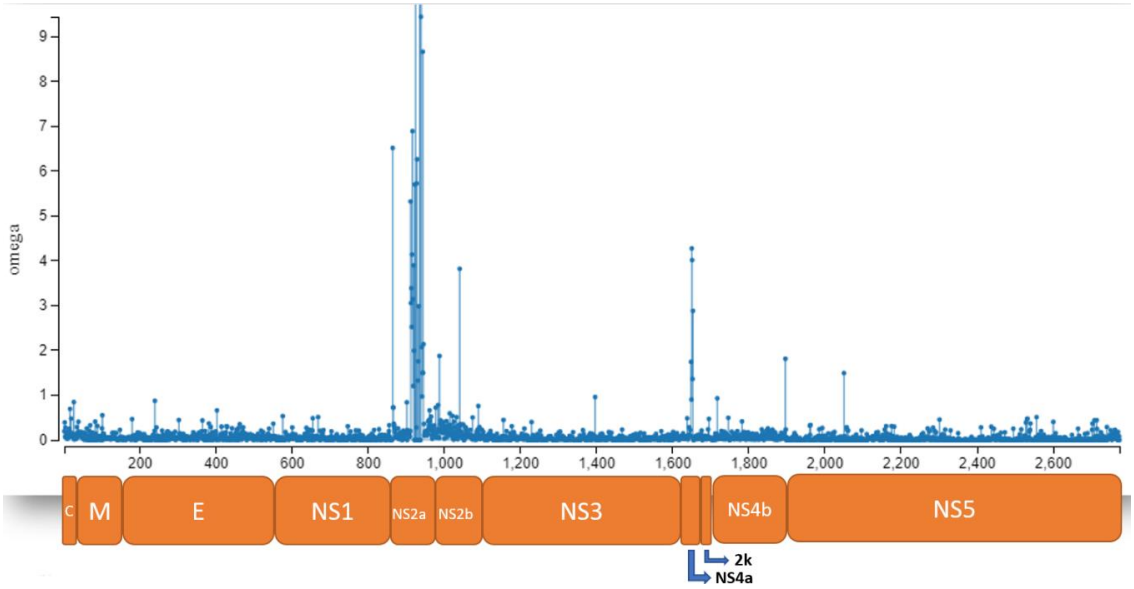


Fig. 1: Assessment of ω values along the cISF complete genome using the FEL function; the different protein coding regions are identified below the x axis, where the indicated coordinates correspond to codon sites in the viral polyprotein coding region.

More specifically, 4 codons were identified by both FEL and SLAC analyses as to be under strong purifying selection. All those codons are located on the NS2a region. Amino acid changes along all sequences for two of these sites (codon number 921 and 929, following numbering on Fig. 1) can be seen in Fig. 2. As described in more detail in the Discussion section in Chapter 4, this *more relaxed* pressure against amino acid change in the NS2a region seems to be important and could be explained but it having little to no effect on RNA synthesis, being mainly involved in the genesis of virus-induced membranes responsible for the transportation of a partially assembled nucleocapsid to the compartment where the final assembly occurs. Compensatory mutations in both NS2A (Leung et al., 2008) and NS3 (NS2a is a cofactor of NS3' activity; Liu et al., 2003) could restore the virion assembly process. Further research should be conducted on this subject.

For mosquito mesoniviruses, while most of their genome is under strong purifying selection, ORF4 had ω values of over 1. However, when the complete genome was evaluated, only 3 sites (none on ORF4) were identified by both FEL and SLAC analysis as to be under strong diversifying selection, one on ORF1a and two on ORF2a. Amino acid changes along all mesonivirus sequences for two of these sites (one for ORF1a and one on ORF2) can be seen in Fig. 3. While purifying selection is commonly found acting under genomes of viruses in the *Nidovirales* order (Ghosh & Chakraborty, 2020; Nam et al., 2019), sites under positive selection on the spike protein can also commonly be identified among these viruses, with the same being observed for mosquito mesonivirus sequences.

For brevihamaparvoviruses, as described in Chapter 6, selective pressure analysis revealed no sites under strong diversifying selection by both FEL and SLAC analysis, which while similar to other parvovirus research (Stamenković et al., 2016), could be hindered by the low number of BHP sequences that have been identified so far.

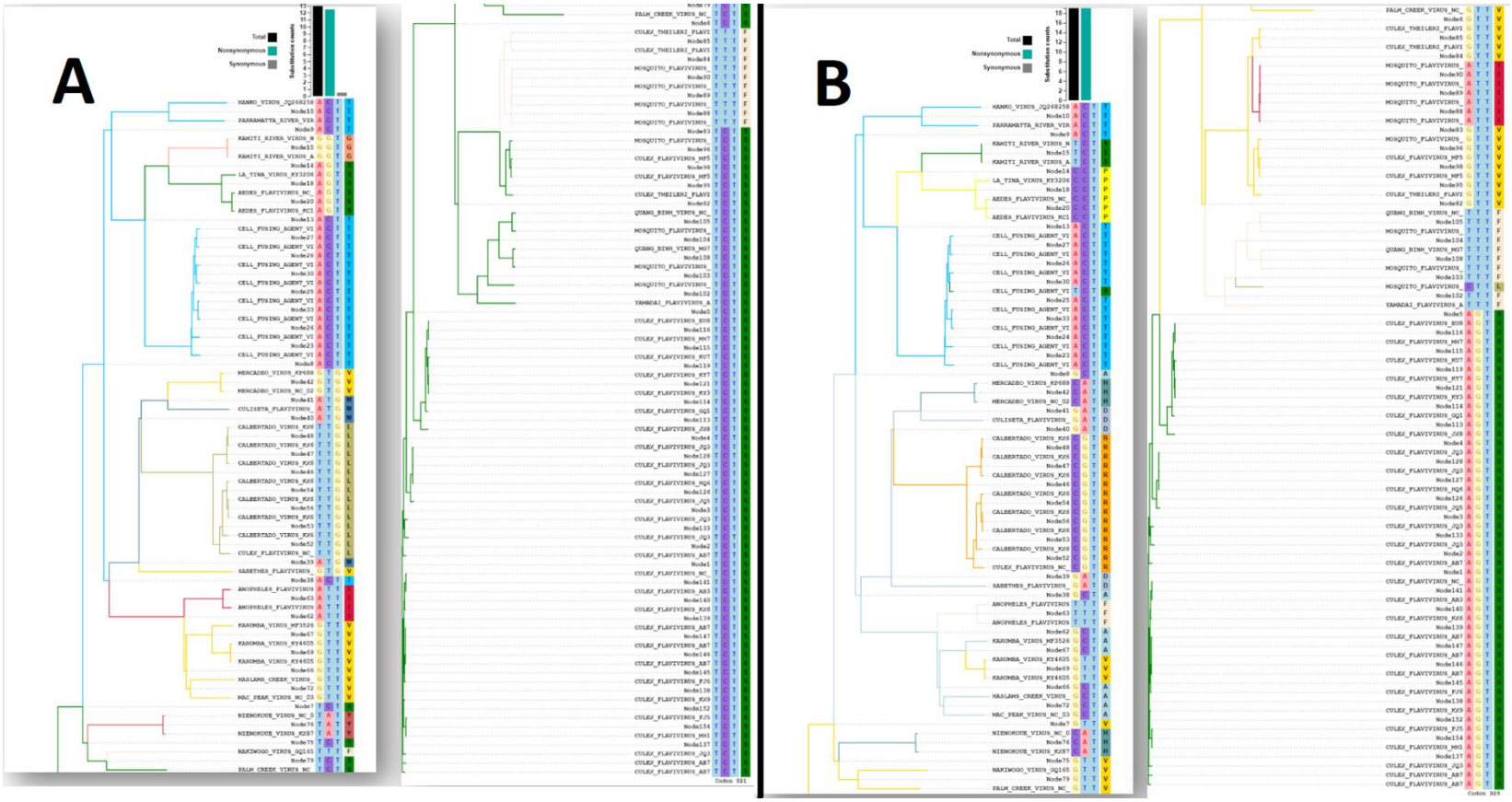


Fig. 2: Amino acid changes between all cISF in codons (A) 921 and (B) 929 (numbering based on Fig.1).

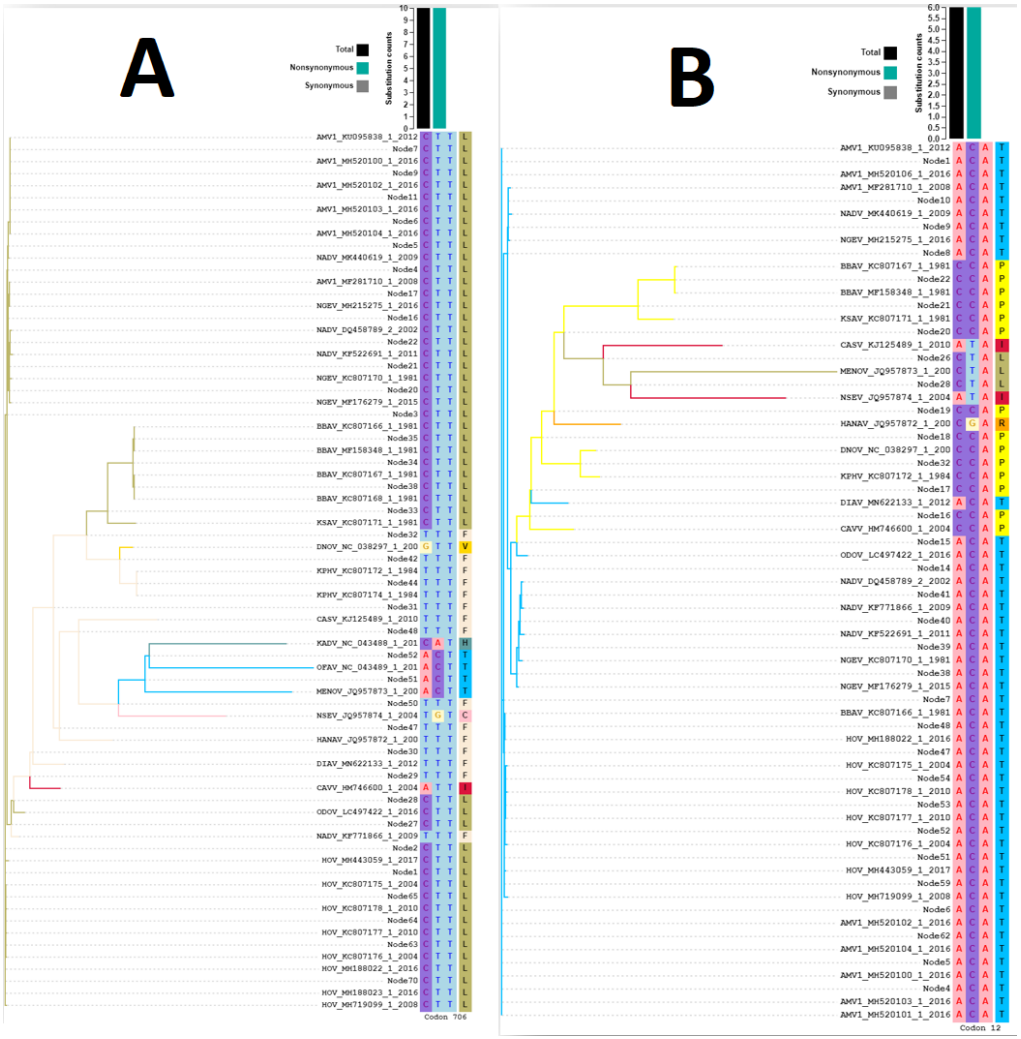


Fig. 3: Amino acid changes between two codons in mosquito mesonivirus sequences, (A) codon 911 – ORF1a and (B) codon 13 – ORF2a, both following Houston virus isolate protein sequence numbering from accession number AYW01750 and AYW01752, respectively

3. Temporal signal analyses

3.1. Introduction

In all research papers included in this thesis, regarding the analysis of spatiotemporal dispersal of different ISVs (Chapter 4, 5 and 6), we first assessed the degree of temporal signal of genomic data included in the analyzed sequence datasets. In this brief section, we will provide more detailed information about the data obtained on temporal signals for all datasets of the three ISV groups analyzed, as well as a brief explanation as to how we optimized the datasets suited to use in phylogeographical analyses (i.e., with high temporal signals).

3.2. Material and Methods

The inspection of the degree of temporal signals from datasets of the three ISV groups analyzed (cISF, mesonivirus, brevihamaparvovirus) was executed as seen in section 2.2 of Chapter 4. Sequences that could negatively impact temporal signals of datasets were removed. We did, however, try to remove as few sequences from our datasets as possible so as to not significantly decrease the final number of sequences analyzed, and possibly jeopardizing phylodynamic analyses.

3.3. Results and Discussion

cISF. Looking back at Table 1 of Chapter 4, two datasets (*Cx. theileri* cISF and CFAV, both for NS5 sequences) had higher R2 values, suggesting they possessed acceptable temporal signal, and would allow spatiotemporal dispersal analysis, unlike all the remainder. Differences between these two datasets and others that were associated with lower divergence-over-time R2 values (using TempEst) can be seen in Fig.4a and 4c. For the Cx. theileri cISF dataset, we removed one sequence that was clearly negatively affecting the analysis, further improving temporal signal data and demonstrating the value of an extensive root-to-tip analysis and removal of outlier sequences (Fig.4b).

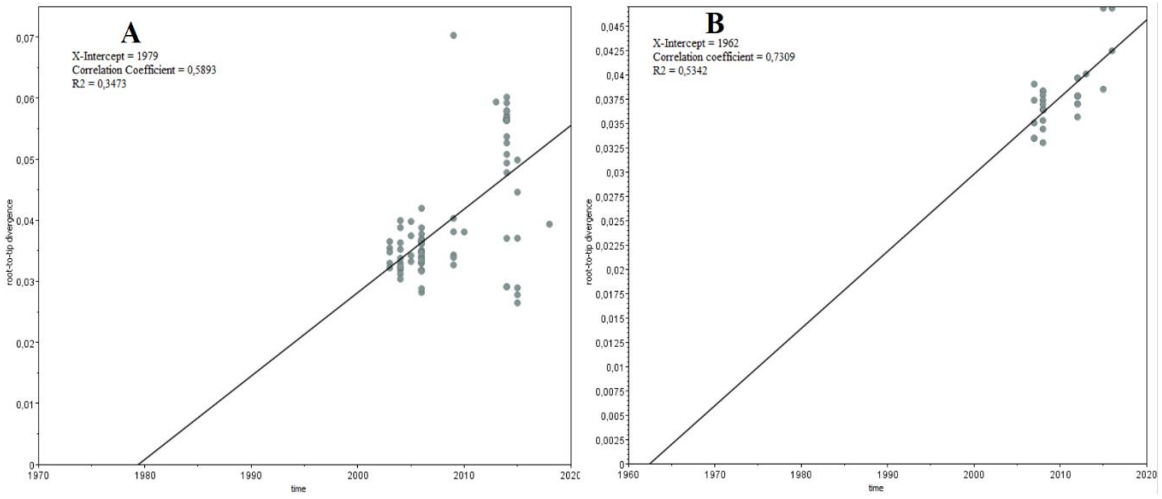


Fig. 4a: Assessment of temporal signal data by TempEst software of (A) *Cx. theileri* cISF NS5 dataset and (B) CFAV NS5 dataset

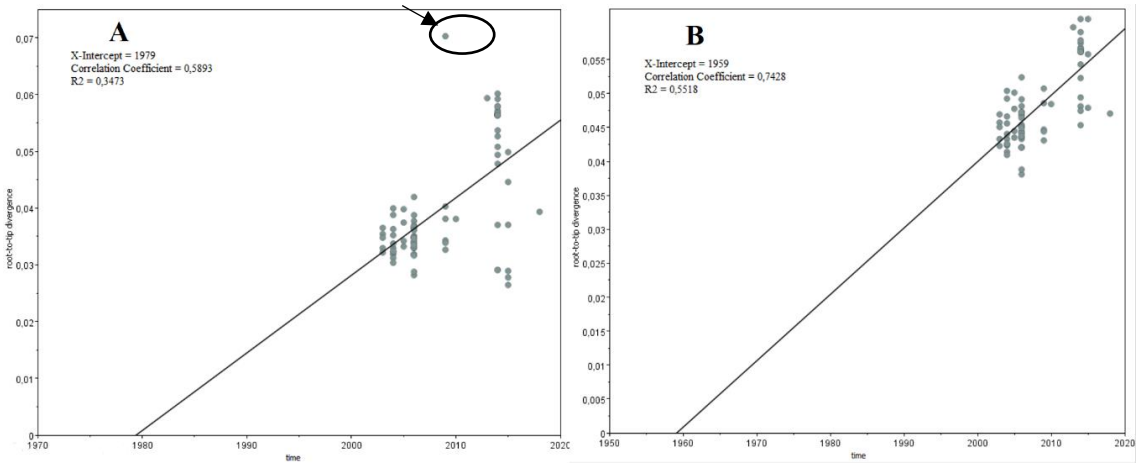


Fig. 4b: (A) Original assessment of temporal signal data of *Cx. theileri* cISF NS5 dataset and (B) differences in values obtained in TempEst by the same analysis after removal of sequence marked in (A).

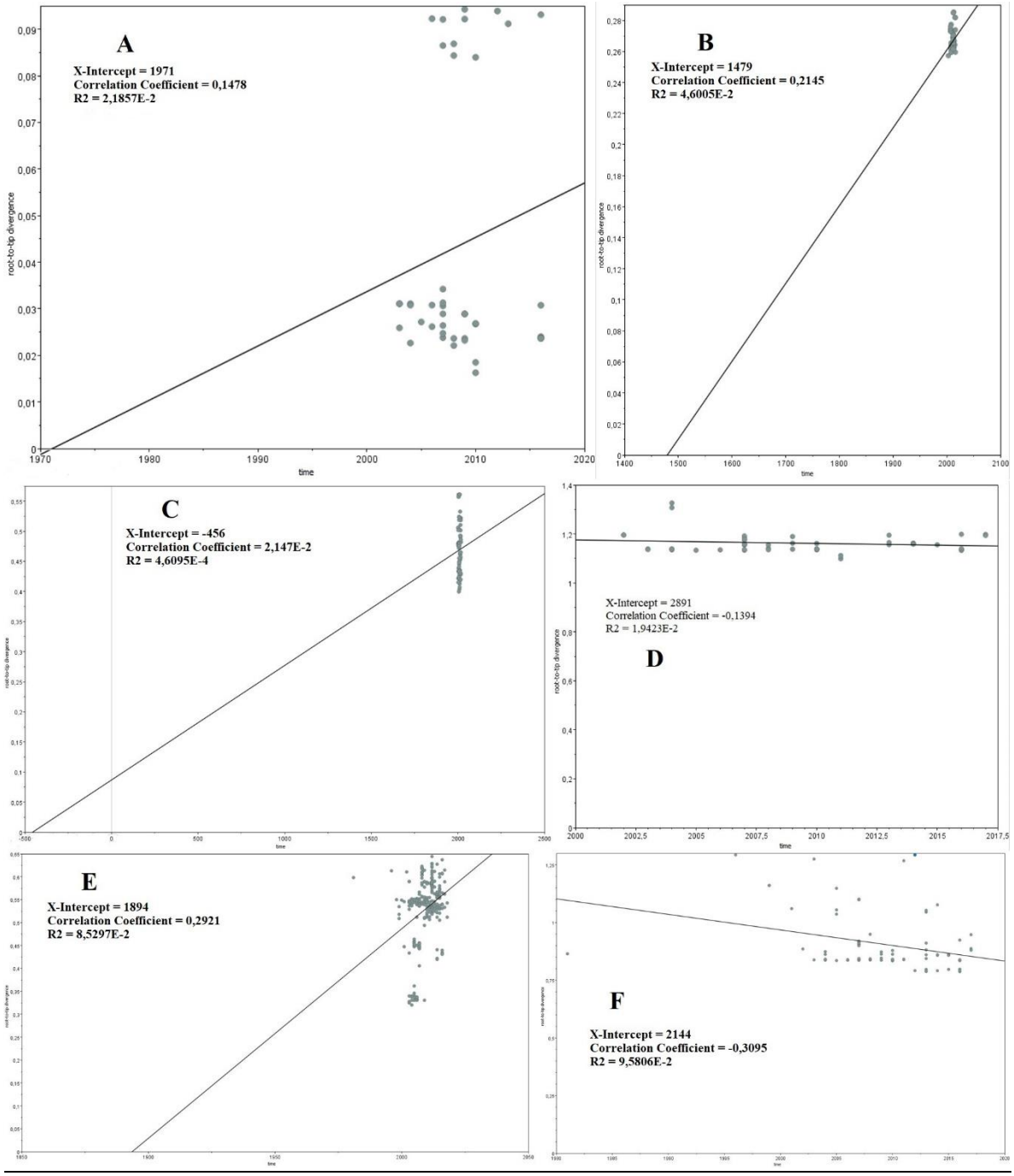


Fig. 4c: Assessment of temporal signal data by TempEst software of (A) *Cx. pipiens* cISF NS5 dataset, (B) *Aedes* cISF NS5 dataset, (C) *Culex* cISF NS5 dataset, (D) *Culex* cISF complete genome dataset, (E) cISF NS5 dataset and (F) cISF complete genome dataset.

Mesonivirus. As depicted in section 3.3 of Chapter 5, significant negative slopes and correlation coefficient values were found for all datasets, hindering potential phylodynamic analyses (Fig. 5). We present here specific examples of obtained TempEst analyses results regarding the analysis of alignments of different sections of the coding region of mosquito mesoniviruses genomic sequences (RdRp and spike proteins) or specific mosquito mesoniviruses belonging only to the AMV1 species (as depicted in Fig.2 of Chapter 5). Similar results were found for all datasets, and this subject should be revisited as new mesonivirus sequences become available in the future.

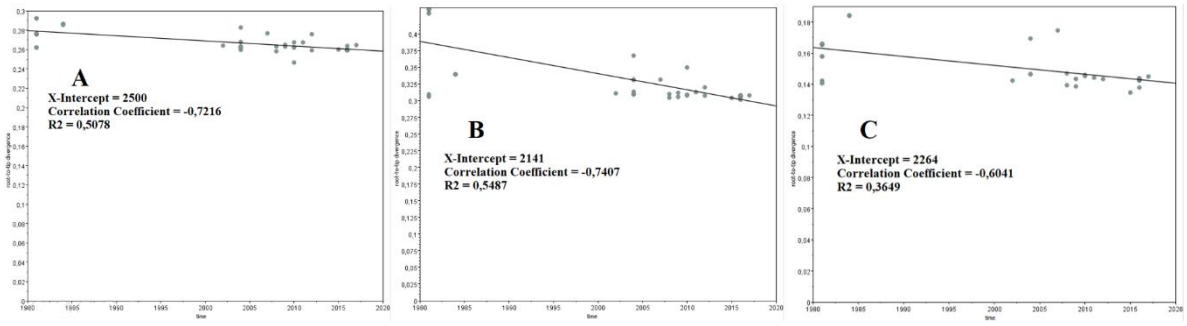


Fig. 5: Assessment of temporal signal data by TempEst software of mosquito mesonivirus (A) RdRp dataset, (B) spike dataset and (C) spike dataset for only mosquito mesonivirus sequences belonging to the AMV1 species.

Brevihamaparvovirus. Three different datasets for three coding regions were considered (NS1, NS2, VP) for temporal signal analysis. According to Table 1 in Chapter 6, the VP coding region provided the more acceptable temporal signal that allowed it to be used as base for a phylodynamic analysis approach. In Fig. 6 we compare the results regarding the VP region to those obtained regarding of the analysis of the NS1 coding region, which displayed the worst results, as far as temporal signal is concerned. Even after removing the four sequences that are clearly set apart from the remainder, instead of improving R^2 value, it actually decreased it, inverting the slope of the regression analysis (from positive to negative) and ultimately deteriorating temporal data available, which means those four sequences could be important to depict genetic diversity over time of brevihamaparvoviruses sequences.

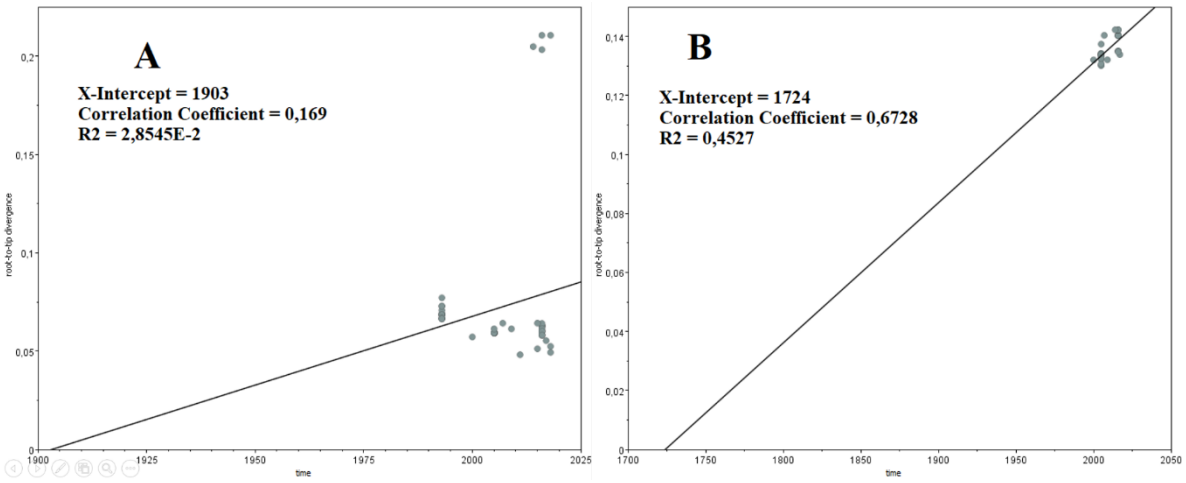


Fig. 6: Assessment of temporal signal data by TempEst software of brevihamaparvovirus (A) NS1 dataset and (B) VP dataset.

4. Spatiotemporal dispersal of cISF

4.1. Introduction

Most arbovirus families are thought to have evolved from insect-only life cycles to on where vertebrate infection becomes an integral part of viral maintenance (Marklewitz et al., 2015). Accordingly, a study on the evolution of flaviviruses suggested that cISFs constitute the ancestral forms from which the vertebrate-infecting flaviviruses have evolved (Shi et al., 2016). However, until recently, no known studies regarding the spatiotemporal dispersion of cISF existed, which could lead to new information regarding their diversity and possible connection to other lineages of flavivirus.

Our extensive study involving the genetic characterization of the most representative genetic lineages found within the cISF branch of the *Flavivirus* phylogenetic tree, as well as a Bayesian-base phylodynamic approach, was conducted and described in Chapter 4. Not only did we find significant differences between distinct sublineages of cISF, we also got evidence of their dispersal from potential MRCAs, which were suggested to have a recent origin (no more than 100 years). However, we could not examine the spatiotemporal dispersal since the oldest tMRCA for all the cISF radiation as described in section 3 of this same chapter, due to low temporal signals obtained. For this reason, we have since attempted to overcome it by analyzing both cISF and other *Flavivirus* lineages (mosquito-borne viruses – MBV, tick-borne viruses – TBV, no-known vector –

NKV and dual-host affiliated insect specific flaviviruses – dhISF) placed altogether in one single dataset, in an attempt to increase the quality of the temporal signal associated with the dataset used, and eventually perform a full phylodynamic analysis of all flavivirus and, more specifically, of all cISF. Using all available cISF nucleotide sequences downloaded from the public databases, as described in Chapter 4, as well as sequences from other *Flavivirus* lineages, we used different bioinformatic tools to perform a Bayesian-based phylodynamic approach and infer the origin and spatiotemporal dispersal of all cISF.

4.2. Material and Methods

As seen in section 3 of this same chapter, partial cISF NS5 sequences had higher temporal signal when compared to complete genome sequences. Following this, we went through all analyses steps as described in section 2 of chapter 4, this time using available partial cISF NS5 sequences (using only one sequence per sublineage/location/year), as well as sequences from other Flavivirus lineages, including all flaviviruses described in Supplementary Fig. 1 in Chapter 4. According to comparative studies with the support of Bayes factor (BF) as predictors using the marginal likelihood values (calculated using both Path-Sampling and Stepping-Stone approaches), our previous study suggested that nonparametric demographic priors were the best candidate models for the phylogeographic study of flavivirus. Accordingly, we investigated and selected the best nonparametric demographic prior (Skygrid), and the geographic spread of flavivirus in continuous space was studied using relaxed random walk (RRW) extensions that model branch-specific variation in dispersal rates (also the best diffusion prior as selected in Chapter 4).

4.3. Results and Discussion

To ensure phylogenetic reconstructions using partial NS5 sequences from all Flavivirus lineages featured acceptable phylogenetic signal, we assessed it by using TREE-PUZZLE. High phylogenetic signal was associated to this specific dataset, with 97.9% of totally resolved quartets that had been randomly sampled. A visual inspection of the

degree of temporal signal of the same dataset was carried using the TempEst software. Final R2 values for partial NS5 genomic sequences of all Flavivirus lineages were much higher when compared to a dataset containing only cISF NS5 sequences (Table 3). As such, we were able to execute a continuous phylogeography analysis using the partial NS5 genomic region dataset of all Flavivirus lineages.

As clearly stated in this thesis, preliminary analysis for selection of best candidate models is vital to phylodynamic approaches. As such, we conducted a similar approach testing three non-parametric coalescent priors along with a Cauchy relaxed random walk (RRW) diffusion model. As seen in Table 4, the non-parametric coalescent prior selected was the Skygrid model.

Table 3: Assessment of temporal signal (Root-to-tip) analyses of partial ns5 coding sequence of (A) only cISF and (B) cISF and other *Flavivirus* lineages, using the TempEst software

	(A) cISF	**(B) All Flavivirus**
Date range	22	91
Slope (rate)	0.0032	0.0011
X-Intercept	1851	1347
Correlation coefficient	0.31	0.52
Root-to-tip analysis (r^2)	0.0096	0.265

Table 4: Evaluation of rates for three different non-parametric priors plus a Cauchy-RRW diffusion approach: analysis of all Flavivirus lineages; bold BF values indicate the best candidate model selected

	root_age [95% HPD]	mean rate [95% HPD]	stdev [95% HPD]	coefficient of variation [95% HPD]	PS1	PS2	SS1	SS2
all Flavi								
GMRF skyride	635 [-124 – 1257]	6,91E-04 [3,69E-04 – 1,03E-03]	1,35E-04 [5,10E-05 – 2,40E-04]	0,19 [0,12 – 0,27]	-42207	-42205	-42206	-42204
Skyline	-113 [-1434 – 923]	5,35E-04 [2,58E-04 – 8,26E-04]	8,87E-05 [2,20E-124 – 1,52E-04]	0,16 [0,09 – 0,25]	-42190	-42182	-42191	-42182
Skygrid	**-58 [-1343 - 946]**	**5,44E-04 [2,81E-04 – 8,30E-04]**	**1,09E-02 [1,84E-03 – 0,03]**	**0,17 [0,09 – 0,24]**	**-42189**	**-42173**	**-42189**	**-42174**

HPD: Highest Probability Density; PS: path sampling; SS: stepping-stone sampling; GMRF: gaussian Markov random field

In Chapter 4, we demonstrated that even if spaciotemporal dispersal analysis is possible in selective subgroups of cISF, when all cISF are taken in consideration, temporal signal analysis is quite poor, which invalidates continuous phylogeography analysis using a Bayesian approach. We suggested the observed low temporal signals associated with most datasets could due to the fact that most cISF sequences had been obtained from biological specimens sampled quite recently, and within a narrow time-range. One possible solution to overcome this limitation would be to execute the same analysis using sequences from all *Flavivirus* lineages, boosting the number of sequences and the data range in our final dataset. Bayesian phylogenetic tree presented in Fig. 7 was reconstructed based on 108 partial NS5 genomic sequences. It was composed of two main robust monophyletic clades, one containing all vertebrate-infecting flaviviruses (including dhISF), with the other containing all cISF, which are split into 3 main robust clades, each corresponding mainly of *Aedes* cISF, *Anopheles* cISF and *Culex* cISF.

The oldest part of the tree splits into two major lineages, one including all cISF and the other the remaining flaviviruses. Curiously, our analysis suggests that their temporal, as well as geographical origins are similar, with African being suggested as the putative birthing place for both lineages. Huge differences do seem to exist in how both clades then disperse over time, with vertebrate-infecting flaviviruses segregating into different sublineages of flaviviruses way sooner than the cISF segregate into its distinct clades. The branches starting from the different ancestors of the cISF lineage appear to be larger than those in the arbovirus-side of the phylogenetic tree, implying that more time has passed between each expansion event. This could, however, result from cISF' reduced sampling size over a narrow time-range. Similar results were found when analyzing the tMRCA of individual cISF sublineages, like the CFAV and *Culex theileri* cISF (the sublineages analyzed in Chapter 4), with similar time frames found for their tMRCA (around 1910-1970).

Indeed, past studies support that some vertebrate-infecting flaviviruses, like dengue and japanese encephalitis virus, evolved from an African ancestral virus over 1000 years in the past (Holmes & Twiddy, 2003; Solomon et al., 2003). Our data suggested that the spatiotemporal dispersal of cISF, also probably from an African origin, with most of its subgroups starting to disperse out of Africa after the 1600's, curiously dates back to the Age of Discovery, where Europeans explored multiple geographic regions across the globe. The dispersal of cISF over time seems to have occurred slowly at first, with most cISF subgroups having a quite recent tMRCA. Such are the cases of CFAV (India, 1971), *Culex theileri* cISF (India, 1940) and *Culex pipiens* cISF (China, 1913). Most tMRCA for cISF monophyletic clades have associated high posterior values for both location and topology. This corroborates our previous work where we analyze the spatiotemporal dispersal of CFAV and *Culex theileri* cISF, with similar results being obtained. After the initial dispersal of cISF in the African continent, different cISF clades seem to have been introduced multiple times into different Asian regions, with the dispersal to other regions happening quite recently. Although we can analyze the time span and geographic locations of most tMRCA with high confidence, curiously most monophyletic clades include sequences from different countries, with some examples of clades with sequences from different continents (like *Culex pipiens* cISF, with sequences from the USA, Japan and even Angola). One interesting find is that two of the sequences obtained in this study

(LC480778 and LC462008 from Portugal and Mozambique, respectively) segregated into one monophyletic clade that recently was composed of only the Nakiwogo virus, with the three sequences possibly comprising a novel species within the genus.

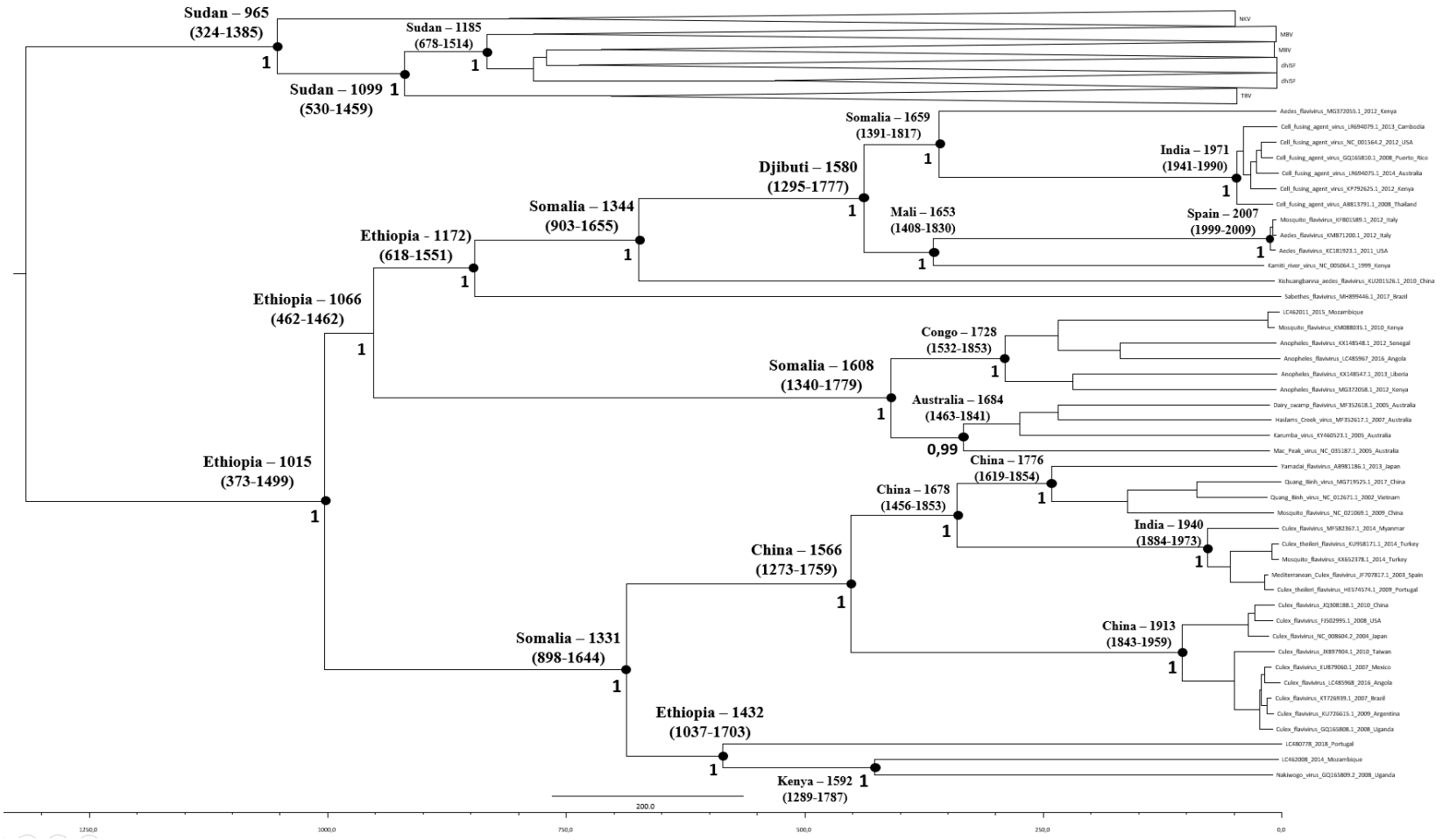


Fig. 7: Continuous phylogeographic analysis of partial cISF *ns5* coding sequence (one sequence per subfamily/location/year) and partial *ns5* coding sequences from other flaviviruses. At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles (only for nodes with geographic origin and/or date of MRCA), while the numbers associated with certain branches indicate the PP for the suggested location.

References

Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., Parreira, R. (2015). Negeviruses found in multiple species of mosquitoes from southern Portugal: Isolation, genetic diversity, and replication in insect cell culture. Virology, 483, 318–328. https://doi.org/10.1016/j.virol.2015.04.021

Ghosh, S., & Chakraborty, S. (2020). Phylogenomics analysis of SARS-COV2 genomes reveals distinct selection pressure on different viral strains. BioMed Research International, 2020, 1–8. https://doi.org/10.1155/2020/5746461

Habbal, W., Monem, F., & Gärtner, B. C. (2009). Comparative evaluation of published cytomegalovirus primers for rapid real-time PCR: Which are the most sensitive? Journal of Medical Microbiology, 58(7), 878–883. https://doi.org/10.1099/jmm.0.010587-0

Holmes, E. C., & Twiddy, S. S. (2003). The origin, emergence and evolutionary genetics of dengue virus. Infection, Genetics and Evolution, 3(1), 19–28. https://doi.org/10.1016/S1567-1348(03)00004-2

Katoh, K., Standley, D.M., 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

Leung, J. Y., Pijlman, G. P., Kondratieva, N., Hyde, J., Mackenzie, J. M., & Khromykh, A. A. (2008). Role of Nonstructural Protein NS2A in Flavivirus Assembly. Journal of Virology, 82(10), 4731–4741. https://doi.org/10.1128/jvi.00002-08

Liu, W. J., Chen, H. B., & Khromykh, A. A. (2003). Molecular and Functional Analyses of Kunjin Virus Infectious cDNA Clones Demonstrate the Essential Roles for NS2A in Virus Assembly and for a Nonconservative Residue in NS3 in RNA Replication. Journal of Virology, 77(14), 7804–7813. https://doi.org/10.1128/jvi.77.14.7804-7813.2003

Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C., & Junglena, S. (2015). Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. Proceedings of the National Academy of Sciences of the

United States of America, 112(24), 7536–7541. https://doi.org/10.1073/pnas.1502036112

Nam, B., Mekuria, Z., Carossino, M., Li, G., Zheng, Y., Zhang, J., Balasuriya, U. B. R. (2019). Intrahost Selection Pressure Drives Equine Arteritis Virus Evolution during Persistent Infection in the Stallion Reproductive Tract. Journal of Virology, 93(12). https://doi.org/10.1128/jvi.00045-19

Pimentel, V., Afonso, R., Nunes, M., Vieira, M. L., Bravo-Barriga, D., Frontera, E., Parreira, R. (2019). Geographic dispersal and genetic diversity of tick-borne phleboviruses (Phenuiviridae, Phlebovirus) as revealed by the analysis of L segment sequences. Ticks and Tick-Borne Diseases, 10(4), 942–948. https://doi.org/10.1016/j.ttbdis.2019.05.001

Shi, M., Lin, X.D., Vasilakis, N., Tian, J.-H., Li, C.-X., Chen, L.J., Zhang, Y.-Z. (2016). Divergent Viruses Discovered in Arthropods and Vertebrates Revise the Evolutionary History of the Flaviviridae and Related Viruses. Journal of Virology, 90(2), 659–669. https://doi.org/10.1128/jvi.02036-15

Solomon, T., Ni, H., Beasley, D. W. C., Ekkelenkamp, M., Cardosa, M. J., & Barrett, A. D. T. (2003). Origin and Evolution of Japanese Encephalitis Virus in Southeast Asia. Journal of Virology, 77(5), 3091–3098. https://doi.org/10.1128/JVI.77.5.3091-3098.2003

Stamenković, G. G., Ćirković, V. S., Šiljić, M. M., Blagojević, J. V., Knežević, A. M., Joksić, I. D., & Stanojević, M. P. (2016). Substitution rate and natural selection in parvovirus B19. Scientific Reports, 6(October), 1–9. https://doi.org/10.1038/srep35759

Yoon, H., & Leitner, T. (2015). PrimerDesign-M: A multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics, 31(9), 1472–1474. https://doi.org/10.1093/bioinformatics/btu832

Chapter 8. Conclusion

1. Concluding remarks and future perspectives

Starting in recent years, our knowledge of the insect virome, especially that of hematophagous arthropods, has been rapidly expanding (Shi et al., 2016). This has been mainly attributed to both advancements in metagenomics, with analysis of virus sequence data obtained directly from a multitude of insect species (Zhang et al., 2018), and genomic analysis technologies, with the introduction of high throughput sequencing as an efficient toolkit to identify both known and emerging insect viruses (Cannon et al., 2021). This combination eventually allows the identification of many viral genomic sequences, especially in the form of RNA. Curiously, many of these sequences have been found integrated into the insect host genome (where they reside as EVEs), which usually remains asymptomatic (Varghese & van Rij, 2018). Some of these viruses were found not to replicate in vertebrate cells (Calisher & Higgs, 2018), and the analyses of the genomes of some of these so-called insect specific viruses were the main focus of this Thesis.

While the identification of the first ISV occurred almost 50 years ago, their characterization, as reported in Chapter 1.2.1., started out slowly with the identification of CFAV in 1975 (Stollar & Thomas, 1975), followed by that of Kamiti river virus in 2003 (Crabtree et al., 2003). A spike in the number of ISVs detected occurred only later, when these viruses started being usually reported either as byproduct of viral surveys aiming the detection of pathogenic arboviruses with global and significant public health impact using degenerate amplification primers (Farfan-Ale et al., 2009), or as the direct product of metagenomics-based viral surveys. New ISV sequences would eventually be obtained from studies carried out across the globe and involving many possible arthropod host species (Nouri et al., 2018), suggesting these viruses do have a widespread geographic distribution and explore a broad range of possible hosts. However, much is still not yet known about ISVs, and our main goal was to provide new information and increase ISV's biodiversity knowledge.

- **New ISV sequences were obtained from multiple geographic regions**

Through the course of this Thesis, we also sought to contribute to the ongoing discovery and characterization of new ISV, via detection of their genomic sequences. We initially proposed to analyse three groups of viruses whose sequences have only been identified in mosquitoes so far: the *Mesoniviridae* family, classical insect-specific flaviviruses

(*Flaviviridae* family) and the *Brevihamaparvovirus* genus (*Parvoviridae* family), as these corresponded to groups of ISVs with an already established widespread distribution and, when at least compared to others, with a significant number of sequences available in the public genomic databases. With this in mind, we also sought to identify ISV sequences from these three specific groups associated with specimens collected in geographical areas where these viruses had not yet been identified. This included the attempted detection of ISVs in mosquitoes collected in Portugal (Chapter 2), Angola (Chapter 3) and Mozambique, using both previously designed primers (targeting the NS5 protein coding-sequence of cISFs; Vázquez et al., 2012) and new ones designed during the course of this project (targeting the RdRp protein coding sequence of mesoniviruses and that of the NS1 and VP proteins of brevihamaparvoviruses). We were able to identify new sequences for all mosquito populations using the proposed methodology, and the new primers were either able to correctly identify new sequences or tested successfully using positive controls. Our study marked the first identification of cISF in mosquitoes from Angola (Chapter 3) and Mozambique (Chapter 4), further identification of cISF in mosquitoes from Portugal (Chapter 2) following previous detection in 2009 by other colleagues in IHMT, as well as the first identification of brevihamaparvoviruses in mosquitoes from Portugal (Chapter 2) and Angola (Chapter 3). This also marked the first time a brevihamaparvovirus sequence was discovered in both Europe and Africa. Not only were we able to expand on the geographical distribution of both cISF and brevihamaparvoviruses, we were also able to identify the first cISF and brevihamaparvovirus associated with *Culiseta* specimens. Unfortunately, no new mesonivirus sequences were identified in all mosquito pools analyzed using the experimental approach undertaken.

As previously described, an ever-expanding number of ISV sequences available could greatly contribute, in the future, to a better understanding of how they evolved and dispersed over time, as well as how they relate and/or what role may they have played in the evolution of pathogenic arboviruses or even of their hosts. While viral surveys of ISVs could not be readily perceived as beneficial outside the scope of the scientific community (since they have little to no direct impacts on public health), attempts should be made to identify and characterize ISVs alongside the more impactful arboviruses. Indeed, sequences available today suggest ISVs could exist in multiple, if not most of, countries

around the world. Even though a mesonivirus sequence was previously identified in a clone obtained by a colleague in IHMT, following a mosquito viral survey in Algarve, suggesting mesoniviruses could be spread in mosquitoes from that region, we did not identify new mesonivirus sequences in mosquitoes collected from Portugal and Angola. Still, mesoniviruses have already been reported in the past in both Europe (Pettersson et al., 2019) and Africa (Diagne et al., 2020), so we could expect to identify them soon as efforts continue to survey their presence around the globe.

- **Vertical transmission suggested for ISV**

While our virus detection strategies did suggest vertical transmission for ISVs (occurring in both males and females collected as immature forms), we could not exclude sexual- or hatchery-associated horizontal transmission since no mosquito eggs were investigated for the amplification of virus sequences. While vertical transmission seems to be the primary mode ISVs explore to ensure their natural maintenance (Vasilakis & Tesh, 2015), rare cases of horizontal transmission have been reported, including one involving the *Rhopalosiphum padi* virus, from the *Dicistroviridae* family (Bonning & Miller, 2010). Future studies should be performed to help clarify how ISVs are transmitted and maintained in nature, which would also be essential when designing future strategies that might explore ISVs against the dispersal of pathogenic arboviruses using previously infected invertebrate vectors.

- **First extensive genetic characterization for these ISV groups**

One may wonder why have such viruses suddenly spiked the interest of the scientific community, which culminated with our intent to specifically investigate more about these three specific groups of ISVs? Two main reasons were considered. First, the potential of ISVs to alter (ideally compromise) the putative vector competence of their insect hosts regarding the transmission of specific pathogenic arboviruses (explored in Chapter 1.2.3.). This topic has been thoroughly addressed and yielded serendipitous optimistic results (Laureti et al., 2020). Multiple data obtained regarding the analysis of different ISVs groups, from their genetic diversity, genome-associated entropy, selective pressure forces acting on their genomes to the analysis and comparison of genome organizations, could help us understand the evolution, natural maintenance strategy as well as the complex interactions these insect viruses establish with their specific hosts. While their

complex vertebrate host-restriction mechanisms have only been explored for a handful of viruses (as depicted in Chapter 1.2.2.), these studies are usually accompanied by systematic reviews of their host range and modes of transmission (Blitvich & Firth, 2015). On the other hand, genomic data analysis of ISVs has been mainly focused on either singular newly-found virus (*Aedes albopictus* brevihamaparvovirus, Chen et al., 2004; Quang Binh virus, Crabtree et al., 2009; Dianke virus, Diagne et al., 2020) or specific subsets of viruses (Culex insect-specific flaviviruses, Bittar et al., 2016). Of the three groups of ISVs we analyzed, which are the most robustly represented at present (i.e., the ones with the highest number of viral sequences available in public genomic databases), only in one particular instance (the *Mesoniviridae* family), has a noteworthy genetic characterization been performed (Vasilakis et al., 2014). However, that study focused more on the analysis of genome sequences and their organization, and only 13 mesoniviruses sequences were available at that time, against the 47 sequences we had available at the start of this PhD project. In the end, we were able to perform a comprehensive and detailed genetic characterization of all three ISV groups in a wider scale.

Second, studies have already demonstrated that insect viruses undergo frequent gene loss or gains, gene change via recombination with gene transfers occurring between viruses, or even involving their permanent establishment in their host genome as EVEs (Shi et al., 2016). With multiple studies strongly suggesting RNA viruses found on insects to be ancestral to those of vertebrates (Marklewitz et al., 2015) and plants (Li et al., 2015), viruses found within insects could have an important role in other viruses' evolutionary history. However, no prior studies for either cISF, brevihamaparvovirus or mesonivirus involving the determination of their ancestral or how they dispersed through time and space had ever been performed. Therefore, it was one of our aims to analyze the dispersion history of specific viral *taxa* through the analysis of their sequences, not before evaluating to what extent their genomes might have undergone genetic exchange via recombination. The information we have managed to produce could help us identify and predict patterns of future dissemination of these viruses and if (and/or how) these viruses are related to pathogenic arboviruses. If, on the one hand, the data we present in this Thesis, regarding the analysis of cISF, mesoniviruses and BHPs, adds new knowledge on

their genetic characteristics, on the other we were also confronted with technical limitations associated with the data available to date.

- **ISV show low diversity and entropy, and mostly evolve under purifying selection**

Interestingly enough, subtle patterns were found in association with the analysis of the three ISV groups. While their genomic organization is similar to other viruses in the same respective family (especially when it comes to ORFs that encode products deemed essential for virus replication), all of them display low genetic diversity, when different regions of the genome scattered throughout its extension are compared. In phylogenetic reconstructions, ISVs cluster into robust monophyletic clades, usually with a well-defined ancestor not directly related to other pathogenic arboviruses. For instance, while all cISF cluster into a single clade, dhISF cluster among mosquito-borne arboviruses. On the other hand, while most genus in the *Parvoviridae* family cluster into multiple monophyletic clusters close to each other, the *Brevihamaparvovirus* genus clusters into a singular clade that is well separated from the remaining parvoviruses. Genetic distance values obtained also support all these claims.

The analysis of selective pressure forces acting upon viral genomes, be them diversifying, purifying or neutral, has been fundamental in investigating genomic diversification (Ghosh & Chakraborty, 2020), and this is especially true in the case of RNA viruses due to their high potential to rapidly accumulate change which grants them with high adaptability (Duffy, 2018). Our analyses indicated that in all three ISV groups, purifying selective pressure has been a major evolutionary drive of virus evolution, in all cases affecting over 99% of all genome sequences analyzed. Strong purifying forces acting on both cISF, mosquito mesonivirus and brevihamaparvovirus genomes suggest that, in general, ISVs seem to be well adapted to their biological niches, to the point where any amino acid relevant changes could eventually disrupt the balance between high viral fitness with little-to-no compromise of that of their hosts. Indeed, no serious cytopathic effects can usually be found when infecting mosquito host cells with most ISVs (Morales-Betoulle et al., 2008; Wang et al., 2017). This overall tendency to stay the same (conservation) is also corroborated by low Shannon entropy levels associated with all three ISV groups. In contrast, positive selection has been key in the evolutionary history

of closely related pathogenic viruses, like the dengue virus (Edgerton et al., 2021) and SARS-CoV-2 (Angeletti et al., 2020).

- **Recombination events appear to not yet be common amongst ISV**

Viral recombination also plays a key role in evolutionary mechanisms of arboviruses, especially when it comes to host-switching, and is a common event found in some species of flaviviruses (Norberg et al., 2013), parvoviruses (Lu et al., 2000) and viruses in the Order *Nidovirales* (Gorbalenya et al., 2006). However, in the course of our analysis only one recombination event was described for cISF and mosquito mesoniviruses, and none were detected for BHP. Strong selective pressure has been suggested to slow down evolution, especially recombinant-driven evolution (Ueda et al., 2017), which might suggest ISVs could be in a kind of "neutral" state when it comes to their single tropism regarding their insect hosts. Host-switching, as defined by their ability to infect/replicate new host cells, has been linked to not only single specific events, be it either the selection of specific mutations or the consequences of recombination events, but to numerous other situations, which would require ISVs to overcome the numerous possible barriers associated with host-restriction (as explored in Chapter 1.2.2.) that ISVs might need to overcome in order to expand their host range, especially in what regards their possible adaptation to the infection of vertebrates. Past studies suggested that dhISF, the ISFs that are mostly related to the flavivirus vertebrate pathogens, had lost their capacity to infect vertebrate cells in the past (Nasar et al., 2012). However, recent studies instead suggested that host range changes from single to multiple tropisms, thanks to the paraphyly of ISFs, happened at different stages of flaviviruses evolutionary history, and that dhISFs could eventually acquire the capacity to become vertebrate pathogens (Öhlund et al., 2019). This theory, while not yet proved, could eventually be applied to cISF as well. The possibility that extant arboviruses such as dengue or Zika might have evolved from single tropism ancestors, restricted to insect hosts in the past, should be evaluated in future studies. Indeed, a recent study suggested that the diversification of flavivirus sequences, especially of the replicase complex (NS3 and NS5), is strongly linked to their current widespread distribution (Caldwell et al., 2022), although more studies are needed to identify which individual proteins have a bigger role on their adaptation to new hosts.

Similar cases of host-switching associated with genetic evolution have been reported in parvoviruses (recent detection of a *Ambidensovirus*, previously associated with insect hosts, in vertebrate hosts such as ducks; Canuti et al., 2021). The identification of meso-like viral sequences (like the one identified on an obligate fungal pathogen, *Leveillula Taurica*) also suggests that host range expansion in the *Mesoniviridae* family could be a stronger possibility than initially thought. Indeed, these meso-like viral sequences had significant differences in genomic organization and were identified in phylogenetic reconstructions not to be closely related neither to mosquito mesoniviruses sequences but also to other virus families in the Order *Nidovirales*. Indeed, they instead form a "sister" clade to the present *Mesoniviridae* (family) monophyletic clade, which indicates these sequences might be regarded as a new genus on the *Mesoniviridae* family or, altogether, even as a new family. Similar cases can be seen, for example, in the *Coronaviridae* family, where coronaviruses with bats and equines as primary hosts share low sequence identity between them (Woo et al., 2009). In summary, our data suggested that since ISVs look to be in an advantageous environment with no major repercussion to their insect hosts (corroborated by evidence of low genetic diversity, strong purifying selection and little to no evidence of gene loss, gain, or recombination events), along with the complexity of host restriction barriers, it is unlikely that sudden host range acquisition will occur, at least if involving only single-site mutations. Instead, it is more plausible that adaptation to new hosts could be driven by multiple substitution events at specific sites, linked to high mutation rates and specific substitution events involving the coding sequence of the VP2 protein, supported the emergence of canine parvoviruses (Shackelton et al., 2005), or by recombination events between closely-related viruses as has been suggested as a key event in the emergence of human coronavirus OC43 (Zhang et al., 2015). Indeed, even though our analysis did not show a pervasive impact of recombination events in the genomic shaping of the virus under analysis, multiple dual-infections have already been reported in nature between ISVs and pathogenic arboviruses (Newman et al., 2011), which could contribute to future genetic exchange between these entities. We also reported high nucleotide substitution rates in all ISV groups, consistent with substitution rates already identified in the past for closely-related viruses. Future studies that specifically identify the key events driving host range expansions could help to predict early instances of host-switching events in different ISV groups.

- **First phylogeographical analysis, albeit with limitations, of multiple ISV groups**

While our phylodynamics analyses had its clear limitations (as expanded on Chapter 4 and 6), including host-associated, geographic, and temporal biases, some insights could be provided and serve as base for similar research in the future. The evaluation of temporal signal and selection of optimal coalescent and demographic dispersal priors for different ISV groups brought out the importance of all these analyses. While insufficient temporal data for mesoniviruses spatiotemporal dispersal analysis were mainly attributed to the poor range of sampling times for existing sequences, we were able to suggest a recent expansion for all BHP sequences and for different sub-lineages in the cISF radiation. Still, the uncertainty of all results obtained, mainly the time estimates of different MRCA and their geographical dispersal were reason for concern. We suggested this was caused primarily by the number of sequences available in public databases (still low, some of them consisting of only partial sequences) and also the temporal biases associated with the recent specimen-collection dates, since ISV detection has only gained traction in the last two decades. Since the number and nature of ISV sequences available could affect genetic diversity and genetic evolution analyses, all these observations should be tested in future research as new sequences become available over time. However, recent research identified sequence diversity as a key factor in the evaluation of temporal signal, with a certain amount of evolutionary change needed to provide acceptable results in a phylodynamic analysis approach (Duchene et al., 2020). Given the low genetic diversity found for all ISV groups, more time and evolutionary change is probably needed before a more accurate assessment can be made of how and for how long these ISV groups have dispersed over time. While we did provide important insight on the execution of phylodynamic analyses and prior selection of these specific ISV groups, following the limitations found at this time, we recommend that all steps, including determination of best performing coalescent and demographic dispersal priors, should be executed again as more sequences are obtained.

While we did suggest that cISF seem to have dispersed more recently when compared to other flaviviruses (Chapter 7), these observations could also be negatively impacted by temporal biases. We suggest a continued monitorization of their expansion conducting similar approaches in the future as new cISF sequences are available. Nonetheless, we

did suggest that the expansion of different flaviviruses occurred through independent evolution of two independent MRCAs, one responsible for the expansion of all cISF sequences available to date, and a completely different one responsible for the expansion of all other flaviviruses, including mosquito and tick-borne arboviruses. This seems to corroborate with previous studies that had suggested that current cISF sequences could correspond to an ancestral lineage of flaviviruses (Cook et al., 2012). While cISF look to have expanded in a similar fashion to the pathogenic arboviruses we know today, our data suggested that these same arboviruses look to have evolved from a different ISF than the ones we analyzed here, one that eventually evolved and adapted to new hosts.

References

Angeletti, S., Benvenuto, D., Bianchi, M., Giovanetti, M., Pascarella, S., & Ciccozzi, M. (2020). COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. Journal of Medical Virology, 92(6), 584–588. https://doi.org/10.1002/jmv.25719

Bittar, C., Machado, D. C., Vedovello, D., Ullmann, L. S., Rahal, P., Araújo Junior, J. P., & Nogueira, M. L. (2016). Genome sequencing and genetic characterization of Culex Flavirirus (CxFV) provides new information about its genotypes. Virology Journal, 13(1), 1–8. https://doi.org/10.1186/s12985-016-0614-3

Blitvich, B. J., & Firth, A. E. (2015). Insect-specific flaviviruses: A systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. Viruses, 7(4), 1927–1959. https://doi.org/10.3390/v7041927

Bonning, B. C., & Miller, W. A. (2010). Dicistroviruses. Annual Review of Entomology, 55(1), 129–150. https://doi.org/10.1146/annurev-ento-112408-085457

Caldwell, H. S., Pata, J. D., & Ciota, A. T. (2022). The Role of the Flavivirus Replicase in Viral Diversity and Adaptation. Viruses, 14(5), 1076. https://doi.org/10.3390/v14051076

Calisher, C. H., & 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(1), 87–103. https://doi.org/10.1146/annurev-ento-020117-043033

Cannon, M. V., Bogale, H. N., Bhalerao, D., Keita, K., Camara, D., Barry, Y., Serre, D. (2021). High-throughput detection of eukaryotic parasites and arboviruses in mosquitoes. Biology Open, 10(7). https://doi.org/10.1242/bio.058855

Canuti, M., Verhoeven, J. T. P., Munro, H. J., Roul, S., Ojkic, D., Robertson, G. J., Whitney, H. G., Dufour, S. C., & Lang, A. S. (2021). Investigating the diversity and host range of novel parvoviruses from north american ducks using epidemiology, phylogenetics, genome structure, and codon usage analysis. Viruses, 13(2), 193. https://doi.org/10.3390/v13020193

Chen, S., Cheng, L., Zhang, Q., Lin, W., Lu, X., Brannan, J., Zhang, J. (2004). Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected Aedes albopictus C6/36 cell line. Virology, 318(1), 123–133. https://doi.org/10.1016/j.virol.2003.09.013

Crabtree, M. B., Sang, R. C., Stollar, V., Dunster, L. M., & Miller, B. R. (2003). Genetic and phenotypic characterization of the newly described insect flavivirus, Kamiti River virus. Archives of Virology, 148(6), 1095–1118. https://doi.org/10.1007/s00705-003-0019-7

Crabtree, M. B., Nga, P. T., & Miller, B. R. (2009). Isolation and characterization of a new mosquito flavivirus, Quang Binh virus, from Vietnam. Archives of Virology, 154(5), 857–860. https://doi.org/10.1007/s00705-009-0373-1

Diagne, M. M., Gaye, A., Ndione, M. H. D., Faye, M., Fall, G., Dieng, I., Sall, A. A. (2020). Dianke virus: A new mesonivirus species isolated from mosquitoes in Eastern Senegal. Virus Research, 275(June 2019), 197802. https://doi.org/10.1016/j.virusres.2019.197802

Duchene, S., Lemey, P., Stadler, T., Ho, S. Y. W., Duchene, D. A., Dhanasekaran, V., & Baele, G. (2020). Bayesian Evaluation of Temporal Signal in Measurably Evolving Populations. Molecular Biology and Evolution, 37(11), 3363–3379. https://doi.org/10.1093/molbev/msaa163

Duffy, S. (2018). Why are RNA virus mutation rates so damn high? PLOS Biology, 16(8), e3000003. https://doi.org/10.1371/journal.pbio.3000003

Edgerton, S. V., Thongsripong, P., Wang, C., Montaya, M., Balmaseda, A., Harris, E., & Bennett, S. N. (2021). Evolution and epidemiologic dynamics of dengue virus in Nicaragua during the emergence of chikungunya and Zika viruses. Infection, Genetics and Evolution, 92, 104680. https://doi.org/10.1016/j.meegid.2020.104680

Farfan-Ale, J. A., Loroño-Pino, M. A., Garcia-Rejon, J. E., Hovav, E., Powers, A. M., Lin, M., Blitvich, B. J. (2009). Detection of RNA from a novel West Nile-like virus and high prevalence of an insect-specific flavivirus in mosquitoes in the Yucatan Peninsula of Mexico. American Journal of Tropical Medicine and Hygiene, 80(1), 85–95. https://doi.org/10.4269/ajtmh.2009.80.85

Ghosh, S., & Chakraborty, S. (2020). Phylogenomics analysis of SARS-COV2 genomes reveals distinct selection pressure on different viral strains. BioMed Research International, 2020, 1–8. https://doi.org/10.1155/2020/5746461

Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J., & Snijder, E. J. (2006). Nidovirales: Evolving the largest RNA virus genome. Virus Research, 117(1), 17–37. https://doi.org/10.1016/j.virusres.2006.01.017

Laureti, M., Paradkar, P. N., Fazakerley, J. K., & Rodriguez-Andres, J. (2020). Superinfection exclusion in mosquitoes and its potential as an arbovirus control strategy. Viruses, 12(11). https://doi.org/10.3390/v12111259

Li, C. X., Shi, M., Tian, J. H., Lin, X. D., Kang, Y. J., Chen, L. J., Zhang, Y. Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. ELife, 2015(4). https://doi.org/10.7554/eLife.05378

Lu, G., Wu, L., Ou, J., & Li, S. (2020). Equine Parvovirus-Hepatitis in China: Characterization of Its Genetic Diversity and Evidence for Natural Recombination Events Between the Chinese and American Strains. Frontiers in Veterinary Science, 7. https://doi.org/10.3389/fvets.2020.00121

Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C., & Junglena, S. (2015). Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. Proceedings of the National Academy of Sciences of the United States of America, 112(24), 7536–7541. https://doi.org/10.1073/pnas.1502036112

Morales-Betoulle, M. E., Pineda, M. L. M., Sosa, S. M., Panella, N., B, M. R. L., Cordón-Rosales, C., Johnson, B. W. (2008). Culex Flavivirus Isolates from Mosquitoes in Guatemala. Journal of Medical Entomology, 45(6), 1187–1190. https://doi.org/10.1093/jmedent/45.6.1187

Nasar, F., Palacios, G., Gorchakov, R. V., Guzman, H., Travassos Da Rosa, A. P., Savji, N., Weaver, S. C. (2012). Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. Proceedings of the National Academy of Sciences of the United States of America, 109(36), 14622–14627. https://doi.org/10.1073/pnas.1204787109

Newman, C. M., Cerutti, F., Anderson, T. K., Hamer, G. L., Walker, E. D., Kitron, U. D., Ruiz, M. O., Brawn, J. D., & Goldberg, T. L. (2011). Culex flavivirus and West Nile virus mosquito coinfection and positive ecological association in Chicago, United States. Vector-Borne and Zoonotic Diseases, 11(8), 1099–1105. https://doi.org/10.1089/vbz.2010.0144

Norberg, P., Roth, A., & Bergström, T. (2013). Genetic recombination of tick-borne flaviviruses among wild-type strains. Virology, 440(2), 105–116. https://doi.org/10.1016/j.virol.2013.02.017

Nouri, S., Matsumura, E. E., Kuo, Y. W., & Falk, B. W. (2018). Insect-specific viruses: from discovery to potential translational applications. Current Opinion in Virology, 33, 33–41. https://doi.org/10.1016/j.coviro.2018.07.006

Öhlund, P., Lundén, H., & Blomström, A. L. (2019). Insect-specific virus evolution and potential effects on vector competence. Virus Genes, 55(2), 127–137. https://doi.org/10.1007/s11262-018-01629-9

Pettersson, J. H. O., Shi, M., Eden, J. S., Holmes, E. C., & 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(11), 1033. https://doi.org/10.3390/v11111033

Shackelton, L. A., Parrish, C. R., Truyen, U., & Holmes, E. C. (2005). High rate of viral evolution associated with the emergence of carnivore parvovirus. Proceedings of the National Academy of Sciences of the United States of America, 102(2), 379–384. https://doi.org/10.1073/pnas.0406765102

Shi, M., Lin, X. D., Vasilakis, N., Tian, J. H., Li, C. X., Chen, L. J., Zhang, Y. Z. (2016). Divergent Viruses Discovered in Arthropods and Vertebrates Revise the Evolutionary History of the Flaviviridae and Related Viruses. Journal of Virology, 90(2), 659–669. https://doi.org/10.1128/jvi.02036-15

Stollar, V., & Thomas, V. L. (1975). An agent in the Aedes aegypti cell line (Peleg) which causes fusion of Aedes albopictus cells. Virology, 64(2), 367–377. https://doi.org/10.1016/0042-6822(75)90113-0

Ueda, M., Takeuchi, N., & Kaneko, K. (2017). Stronger selection can slow down evolution driven by recombination on a smooth fitness landscape. PLoS ONE, 12(8), e0183120. https://doi.org/10.1371/journal.pone.0183120

Varghese, F. S., & van Rij, R. P. (2018). Insect virus discovery by metagenomic and cell culture-based approaches. In Methods in Molecular Biology (Vol. 1746, pp. 197–213). https://doi.org/10.1007/978-1-4939-7683-6_16

Vasilakis, N., Guzman, H., Firth, C., Forrester, N. L., Widen, S. G., Wood, T. G., Rossi, S. L., Ghedin, E., Popov, V., Blasdell, K. R., Walker, P. J., & Tesh, R. B. (2014). Mesoniviruses are mosquito-specific viruses with extensive geographic distribution and host range. Virology Journal, 11(1), 1–12. https://doi.org/10.1186/1743-422X-11-9715

Vasilakis, N., & Tesh, R. B. (2015). Insect-specific viruses and their potential impact on arbovirus transmission. Current Opinion in Virology, 15, 69–74. https://doi.org/10.1016/j.coviro.2015.08.007

Vázquez, A., Sánchez-Seco, M.-P., Palacios, G., Molero, F., Reyes, N., Ruiz, S., Tenorio, A. (2012). Novel Flaviviruses Detected in Different Species of Mosquitoes in

Spain. Vector-Borne and Zoonotic Diseases, 12(3), 223–229. https://doi.org/10.1089/vbz.2011.0687

Wang, Y., Xia, H., Zhang, B., Liu, X., & Yuan, Z. (2017). Isolation and characterization of a novel mesonivirus from Culex mosquitoes in China. Virus Research, 240, 130–139. https://doi.org/10.1016/j.virusres.2017.08.001

Woo, P. C. Y., Lau, S. K. P., Huang, Y., & Yuen, K. Y. (2009). Coronavirus diversity, phylogeny and interspecies jumping. Experimental Biology and Medicine, 234(10), 1117–1127. https://doi.org/10.3181/0903-MR-94

Zhang, Y., Li, J., Xiao, Y., Zhang, J., Wang, Y., Chen, L., Paranhos-Baccalà, G., Ren, L., & Wang, J. (2015). Genotype shift in human coronavirus OC43 and emergence of a novel genotype by natural recombination. Journal of Infection, 70(6), 641–650. https://doi.org/10.1016/j.jinf.2014.12.005

Zhang, Y. Z., Shi, M., & Holmes, E. C. (2018). Using Metagenomics to Characterize an Expanding Virosphere. Cell, 172(6), 1168–1172. https://doi.org/10.1016/j.cell.2018.02.043

Annex

Excerpts from methods referenced on Chapter 2 and 3, published as:

Carapeta, S., do Bem, B., McGuinness, J., Esteves, A., Abecasis, A., Lopes, Â., de Matos, A.P., Piedade, J., de Almeida, A.P.G., Parreira, R., 2015. Negeviruses found in multiple species of mosquitoes from southern Portugal: isolation, genetic diversity, and replication in insect cell culture. Virology 483, 318–328. https://doi.org/10.1016/j.virol.2015.04.021

"Mosquito collection and homogenate preparation

Mosquito homogenates were prepared by mechanical disruption of adult specimens using glass beads, as previously described (Huang et al., 2001), using 1.3 ml of phosphate buffered saline (PBS) supplemented with 4% of Bovine Serum Albumin (Fraction V; NZYtech, Lisbon, Portugal). After clarification of the mosquito macerates by centrifugation at 13,000 x g (4ºC for 10 min), supernatants were kept at -80ºC until further use.

Cell culture, and virus isolation, titration, and purification

The C6/36 cell line, established from macerates of Aedes albopictus larvae, was used for virus isolation. Cells were maintained at 28çC (in the absence of CO2) in Leibovitz's L-15 medium (Lonza, Walkersville, MD, USA) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) (Lonza, Walkersville, MD, USA), 2 mM L-glutamine (Gibco BRL, Gaithersburg, MD, USA), 100 U/mL penicillin (Gibco BRL, Gaithersburg, MD, USA), 100 µg/mL streptomycin (Gibco BRL, Gaithersburg, MD, USA), and 0.26% (v/v) triptose phosphate broth (AppliChem GmbH, Darmstadt, Germany).

Aliquots of clarified mosquito homogenates (approximately 500 µl) were sterilized through 0.22 µm disposable PVDF filters (Millex-GV, Millipore Corp., Bedford, MA, USA), diluted in an equal volume of PBS, and used to inoculate subconfluent C6/36 cell monolayers in T25 flasks (Thermo Scientific Nunc, Roskilde, Denmark). After 1 h at room temperature (allowing for viral adsorption), the inoculum was removed, Leibovitz's L-15 medium (5% FBS) added, and the cultures incubated at 28 1C for up to a week. Culture supernatants were collected, after a second- or third-blind passage, depending on the magnitude of the cytopathic effect (CPE) observed, and stored at -80ºC.

Separation of OCFVPT from OCNV (strain 174) (henceforth designated OCNV; see nomenclature description in the Results section) was carried out by limiting dilution, starting from a cell supernatant in which both viruses were present at different titers. This virus stock was serially-diluted, and each dilution used to infect C6/36 cells (as previously indicated) up to the point where the genome of OCNV, but not that of OCFVPT, could still be detected in the culture supernatant, by RT-PCR, using NegIF/NegIR and AcFV11F/AcFV21R primer pairs (Ferreira et al., 2013), respectively. For plaque assay titration, monolayers of C6/36 cells were inoculated with serial dilutions of virus samples in PBS. After adsorption for 1 h, cells were covered with 2% Sephadexs G-50 (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) in Leibovitz's L-15 medium with 2% FBS, and incubated at 28ºC for 48 h. Cells were fixed with 4% formaldehyde in PBS and stained with 0.1% crystal violet and 1% methanol in PBS, before plaque counting.

Culture supernatants from C6/36 cells harvested 48 h post-infection (p.i.) with OCNV were clarified by centrifugation at 2000xg for 10 min and the virus was precipitated overnight at 4ºC in the presence of 7% PEG6000 and 2.3% NaCl. The viral particles were collected by centrifugation (4000xg, 30 min, at 4ºC), resuspended in TEN buffer (50 mM Tris–HCl pH 7.4, 100 mM NaCl, 1 mM EDTA), loaded onto a discontinuous 20–70% sucrose gradient, and centrifuged for 1 h at 270,000xg. The virus was collected from the sucrose interface and subsequently loaded onto an Amicon Ultra centrifugal filter (Merck Millipore, Billerica, MA, USA) with a 100 kDa cutoff, and centrifuged at 4000xg, at 4ºC, for 15 min.

Nucleic acid extraction, purification, amplification, and DNA sequencing

Viral RNA was extracted from 150 ml of either clarified mosquito macerate or culture supernatant, using the ZR Viral RNA Kit™ (Zymo Research, Irvine, CA, USA) according to the manufacturer's recommendations. Reverse transcription of viral RNA was carried out with the Phusion RT-PCR Kit (Thermo Fisher Scientific, Waltham, MA, USA) using random hexaprimers and 5 ml of the RNA extract. The obtained cDNA served as template for the amplification of viral sequences using Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific, Waltham, MA, USA) or NZYTaq 2x Green Master Mix (NZYTech, Lisbon, Portugal), and the oligonucleotides listed in Supplementary Table 1. DNA amplicons were purified with the DNA Clean & Concentrator™-5 (Zymo Research, Irvine, CA, USA) and directly sequenced. The completion of the viral genomic sequence was carried out by Rapid Amplification of cDNA Ends (RACE) essentially as described by Tillett et al., (2000), with modifications suggested by Li et al., (2005). The sequence of the 5'-end was determined after amplification of viral cDNA using the DT89/R2 and DT89/R0 primer pairs (Supplementary Table 1) in the 1st and 2nd rounds of PCR, respectively. Similarly, the 3'-end of the viral genome sequence was completed after amplification of viral cDNA using the DT89/F9 (1st round) and DT89/raceF (2nd round) pairs of primers (Supplementary Table 1).

To study the kinetics of viral RNA synthesis, selective strand-specific amplification of viral RNA was carried out as previously described (Plaskon et al., 2009) on C6/36 cells subjected to cold-synchronized infection. Virus adsorption was allowed to occur at 4ºC for 1h (m.o.i >410), was followed by thorough wash of cell monolayers with cold PBS, before shifting to 28 1C for viral replication to occur. Tagged-primers were used to decrease the chances of false-priming of RNA during RT reactions, which may occur in the absence of any specific oligonucleotide (Peyrefitte et al., 2003). These primers, to which a 21-nucleotide (nt.) 50-tag (GGCCGTCATGGTGGCGAATAA) with no known homology to negevirus sequences was added, are listed in Supplementary Table 1. The detection of either the (+) or (-) viral strands involved the preparation of strand-specific cDNA using the tagF174 (- strand) or tagR174 (+ strand) primers, followed by 35 cycles of conventional PCR amplification (94ºC, 30s; 57ºC, 30s; 72ºC, 45s) using the Tag/R174

(+ strand) and Tag/F174 (- strand) primer pairs. Alternatively, PCR reactions were carried out in real-time format in a Rotor-Gene 3000 thermocycler (Corbett Research, St. Neots, UK), in reaction volumes of 25 µl, using the Maxima SYBR Green qPCR Master Mix (2X) (Thermo Fisher Scientific, Waltham, MA, USA) and 400 nM of each primer, allowing for the detection of amplicon synthesis as a function of fluorescence emission by DNA-bound SYBR Green I. Positive amplification results were defined as those for which fluorescence intensity increased exponentially over, at least, five consecutive cycles, with a cycle threshold (Ct) <30."

Excerpts from methods referenced on Chapter 2 and Chapter 3, published as:

Vázquez, A., Sánchez-Seco, M.-P., Palacios, G., Molero, F., Reyes, N., Ruiz, S., Aranda, C., Marqués, E., Escosa, R., Moreno, J., Figuerola, J., Tenorio, A., 2012. Novel flaviviruses detected in different species of mosquitoes in Spain. Vector-Borne Zoonotic Dis. 12, 223–229. https://doi.org/10.1089/vbz.2011.0687

"Generic NS5 RT-nested-PCR

Nucleotide sequences of complete NS5 genes of different flaviviruses were obtained from GenBank (National Institute of Health, Bethesda, MD) and aligned by using the algorithm Clustal X as implemented in the MEGA 4.0 software (Tamura et al. 2007). Degenerated primers were designed based on conserved motifs of the NS5 gene; primers selected were 1NS5F: 5'$_{9035}$-GCATCTAYAWCAYNATGGG-$_{9053}$3', 1NS5Re: 5'$_{10129}$-CCANACNYNRTTCCANAC-$_{10146}$3', 2NS5F: 5'$_{9103}$-GCNATNTGGTWYATGTGG-$_{9120}$3' and 2NS5Re: 5'$_{10103}$- CATRTCTTCNGTNGTCATCC-$_{10122}$3'. Indicated positions correspond to the sequence of WNV strain NY99-flamingo382–99 (accession number: AF196835).

Viral RNA was extracted from mosquito pools or cell culture supernatants by using a QIAamp Viral RNA Mini Kit (QIAGEN). RT-PCR was conducted by using One-Step RTPCR kit (QIAGEN) using degenerated primer set 1NS5F/1NS5Re. First amplification

profile was 50ºC for 45 min and 95ºC 15 min, followed by 40 cycles of 94ºC for 1 min, 50ºC for 4 min, and 72ºC for 1 min, with a final extension for 10 min at 72ºC. Second amplification was carried out in a final volume of 50 µL and contained 5mM $MgCl_2$ (Perkin Elmer-Cetus), 0.1mM of each dNTP (Amersham Pharmacia Biotech), 60 pmol of each primer, 2.5U of AmpliTaq DNA Polymerase (Applied Biosystems), and 1 µL of the first amplification product. Second amplification profile was 94ºC for 5 min, followed by 40 cycles of 94ºC for 1 min, 50ºC for 3 min, and 72ºC for 1 min, with a final extension for 10 min at 72ºC. The reactions were performed in a Peltier Thermal Cycler (PTC-200; MJ Research, Watertown). The amplified products were visualized by ethidium bromide staining after electrophoresis on a 1.5% high-resolution agarose gel (MS8; Hispanlab).”

Excerpts from methods referenced on Chapter 2 and Chapter 3, published as:

Parreira, R., Cook, S., Lopes, Â., de Matos, A.P., de Almeida, A.P.G., Piedade, J., Esteves, A., 2012. Genetic characterization of an insect-specific flavivirus isolated from Culex theileri mosquitoes collected in southern Portugal. Virus Res. 167, 152–161. https://doi.org/10.1016/j.virusres.2012.04.010

“Partial mitochondrial cytochrome c oxidase subunit I (COI) sequences were amplified from total genomic DNA, extracted from mosquito homogenates with the ZymoBeadTM Genomic DNA kit (Zymo Research, Irvine, CA), and the PuRe Taq Ready-to-Go PCR Beads (GE Healthcare, Dornstadt, Germany), using primers and reaction conditions previously described (Cook et al., 2009; described by Folmer et al., 1994). The obtained amplicons were purified, cloned in pGEM®T Easy (Promega, Madison, WI), and sequenced.”

Primers and reaction conditions described in Parreira et al., 2012, as published in:

Cook, S., Moureau, G., Harbach, R. E., Mukwaya, L., Goodger, K., Ssenfuka, F., Gould, E., Holmes, E. C., & de Lamballerie, X. (2009). Isolation of a novel species of flavivirus and a new strain of Culex flavivirus (Flaviviridae) from a natural mosquito population in Uganda. Journal of General Virology, 90(11), 2669–2678. https://doi.org/10.1099/vir.0.014183-0

“The COI gene was amplified by using primers UEA3 (59-TATRGCWTTYCCWCGAATAAATAA-39) (Lunt et al., 1996) and Fly10 (59-ASTGCACTAATCTGCCATATTAG-39) (Sallum et al., 2002) according to Cook et al. (2006) (hereafter referred to as the ‘Fly’ region). An additional overlapping region, the ‘barcode’ section of the COI gene, was amplified using the primers LCO1490 (59-GGTCAACAAATCATAAAGATATTGG-39) and HCO2198 (59-TAAACTTCAGGGTGACCAAAAAATCA-39) [both from Folmer et al. (1994)] with a primer concentration of 10 mM and reaction conditions of 5 min at 95 uC; 40 cycles of 30 s at 95 uC, 30 s at 48 uC and 45 s at 72 uC; followed by a final extension time of 5 min at 72 uC.”