VIRUS EVOLUTION

Virus Evolution, 2022, **8(1)**, 1–15

DOI: https://doi.org/10.1093/ve/veac036

Advance access publication date: 16 April 2022

Research Article

Defining virus-carrier networks that shape the composition of the mosquito core virome of a local ecosystem

Konstantinos Konstantinidis,[1,†,‡] Nikolas Dovrolis,[1,†,§] Adamantia Kouvela,[1] Katerina Kassela,[1] Maria Goreti Rosa Freitas,[2] Andreas Nearchou,[1] Michael de Courcy Williams,[1] Stavroula Veletza,[1] and Ioannis Karakasiliotis[1,*,**]

[1]Department of Medicine, Laboratory of Biology, Democritus University of Thrace, Dragana, Alexandroupolis, PC 68100, Greece and [2]Laboratório de Mosquitoes Transmissores de Hematozoários, Instituto Oswaldo Cruz-Fiocruz, Pavilhão Carlos Chagas, Avenida Brasil, 4365 - Manguinhos, Rio de Janeiro, PC 21040-900, Brazil

[†]Authors contributed equally to the work.

[‡]https://orcid.org/0000-0002-9093-758X

[§]https://orcid.org/0000-0003-3832-9939

[**]https://orcid.org/0000-0003-3004-8104

*Corresponding author: E-mail: ioakarak@med.duth.gr

Abstract

Mosquitoes are the most important vectors of emerging infectious diseases. During the past decade, our understanding of the diversity of viruses they carry has greatly expanded. Most of these viruses are considered mosquito-specific, but there is increasing evidence that these viruses may affect the vector competence of mosquitoes. Metagenomics approaches have focused on specific mosquito species for the identification of what is called the core virome. Despite the fact that, in most ecosystems, multiple species may participate in virus emergence and circulation, there is a lack of understanding of the virus-carrier/host network for both vector-borne and mosquito-specific viruses. Here, we studied the core virome of mosquitoes in a diverse local ecosystem that had 24 different mosquito species. The analysis of the viromes of these 24 mosquito species resulted in the identification of 34 viruses, which included 15 novel viruses, as determined according to the species demarcation criteria of the respective virus families. Most of the mosquito species had never been analysed previously, and a comparison of the individual viromes of the 24 mosquito species revealed novel relationships among mosquito species and virus families. Groups of related viruses and mosquito species from multiple genera formed a complex web in the local ecosystem. Furthermore, analyses of the virome of mixed-species pools of mosquitoes from representative traps of the local ecosystem showed almost complete overlap with the individual-species viromes identified in the study. Quantitative analysis of viruses' relative abundance revealed a linear relationship to the abundance of the respective carrier/host mosquito species, supporting the theory of a stable core virome in the most abundant species of the local ecosystem. Finally, our study highlights the importance of using a holistic approach to investigating mosquito viromes relationships in rich and diverse ecosystems.

Key words: mosquitoes; virome; ecosystem; metagenomics; vectors.

1. Background

Arboviruses pose a significant threat to public health worldwide by causing epidemics such as Dengue, West Nile, Zika, and chikungunya diseases with considerable repercussions for the infected vertebrate hosts (Mayer, Tesh, and Vasilakis 2017). Undoubtedly, tropical, rural and agricultural areas are plagued by arboviruses, while climate change, urbanization, and globalization affect the geographic and seasonal patterns of vector-borne diseases, facilitating the efficient reproduction of insects and the pathogens they may carry (Elbers, Koenraadt, and Meiswinkel 2015; Rocklöv and Dubrow 2020). Mosquitoes are major vectors for most known vector-borne diseases. Global surveillance data have reported a considerable increase in mosquito-borne diseases, and pest control of mosquitoes is a top priority during their activity period to prevent future outbreaks (Dahmana and Mediannikov 2020; Rocklöv and Dubrow 2020). In parallel to important human and other animal pathogens, mosquitoes carry a significant number of mosquito-specific viruses that either are limited to the insect host or no other secondary host has been identified to date (Roundy et al. 2017; Agboli et al. 2019). Viruses that are considered to be insect-specific are often transmitted vertically (Brito et al. 2021; C. Shi et al. 2020), through generations, and are recently of great interest as modulators of mosquito vector competency (Öhlund, Lundén, and Blomström 2019; Pyke et al. 2021; Viglietta et al. 2021).

Significant efforts have been made to elucidate the virome of insects in endeavours to characterize the constituents of the 'virosphere' of natural habitats and identify sources where infectious

agents may emerge. Two key publications detailed the virosphere of several provinces in China. Li and colleagues analysed 70 species from four different arthropod classes through RNA sequencing, discovering 112 novel viruses in China (C.-X.; Li et al. 2015). Shi and colleagues profiled the transcriptomes of over 220 invertebrate species sampled across nine animal phyla and reported the discovery of 1,445 RNA viruses within the same provinces of China, filling virus phylogeny gaps (M. Shi et al. 2016). As mosquitoes are a top priority, considering their vector potential, several studies have focused on metagenomics approaches to defining the mosquito virome in pools of mosquitoes that contain either multiple or single species. Mosquito species, usually belonging to *Aedes (Ae.)*, *Anopheles* (*An.*), and *Culex (Cx.)* genera that serve as important vectors worldwide have been analysed in the past for the presence of viruses. The vast majority of the identified viruses are RNA viruses with segmented or non-segmented RNA genomes.

Comparative analysis of different mosquito species in the same habitat has highlighted viromes with distinct compositions (C. Shi et al. 2019). Mosquito species that belong to the same genus present less distinctive viromes (M. Shi et al. 2017), although in-depth comparison has identified virome patterns that differentiate related species (Pettersson et al. 2019). The idea of a core virome follows that of the commensal (core) microbiome that, in most cases, does not give rise to clinical symptoms but is rather considered symbiotic (Haynes and Rohwer 2011; Neu, Allen, and Roy 2021). The viruses of the core virome, as opposed to the viruses that cause acute symptomatic infections, have been characterized in a variety of organisms, including humans (Haynes and Rohwer 2011), animals (de Brito et al. 2021), plants (Fetters et al. 2022), and fungi (Jia et al. 2021). Core viromes have been shown to be significantly similar (stable), in terms of virus species occurrence, among individuals of same mosquito species while being significantly diverse amongst different species (C. Shi et al. 2019, 2020). Interestingly, the core virome remained stable in laboratory-grown and field-collected *Ae. albopictus* across developmental stages, which is consistent with the vertical transmission of such viruses (C. Shi et al. 2020).

Comparison of core viromes among mosquito species in a habitat is a first step in defining the ecological dynamics of mosquito-associated viruses that may shape the ecology and transmission of pathogens, particularly those of human or animal health importance such as West Nile virus (Hobson-Peters et al. 2013; Vasilakis and Tesh 2015; Hall-Mendelin et al. 2016). Such comparisons are usually restricted to between two or just a few mosquito species within a particular ecosystem and normally target medically important species of the region (Fauver et al. 2016; Pettersson et al. 2019; C. Shi et al. 2019). A meta-analysis of mosquito-associated viruses in China, with the aim of summarizing the distribution and diversity of these viruses in the country's provinces, recorded the wide distribution of numerous virus families among seven mosquito genera (Atoni et al. 2020). The distribution of similar viruses among different mosquito species highlights the potential importance of all mosquito species that occur in an ecosystem in determining the overall circulation of mosquito-associated viruses. In the present report, we aim to determine the core virome of all mosquito species present in an ecosystem and analyse the viruses they carry/host. Comparative analysis of the most abundant viruses present in the ecosystem allowed us to identify the distinct signatures of viruses in the different mosquito species and determine those species that may act as virus transmission bridges between different genera.

2. Materials and methods

2.1 Mosquito collection and identification

Adult mosquitoes were field-collected using Center of Disease Control (CDC) light traps baited with CO_2. Collection points spanned the region of Eastern Macedonia and Thrace in northern Greece (Fig. 1, Supplementary Table S1) during two different periods of intensified mosquito activity: April–October 2018 and April–October 2019. The region of Eastern Macedonia and Thrace was subdivided into five discrete subareas as defined by the five major municipalities of Drama, Xanthi, Komotini, Alexandroupoli, and Orestiada. Traps from the collection points of each subarea were analysed in terms of species identification and abundance. After collection, the samples were stored and delivered to the laboratory on dry ice. Mosquito specimens were examined over a bed of crushed ice to maintain their condition at all times, both during sample sorting and species identification. Good-quality intact individuals were stored at –80°C prior to RNA and DNA extraction. Female mosquitoes were identified using external morphological features. Species nomenclature follows Harbach (2018) and generic abbreviations follow Wilkerson et al. (2015). Morphological identification was done using a combination of the keys of Samanidou-Voyadjoglou (2001) and Samanidou-Voyadjoglou and Harbach (2011) and the online resource MosKey-Tool (Gunay, Picard, and Robert 2016). For the genus *Anopheles*, a key to females (Glick 1992) provided additional morphological characters relevant to the Greek fauna.

2.2 Species identification through cytochrome c oxidase subunit I barcoding

Mosquito individuals were homogenized using a pellet pestle (Eppendorf) and total RNA was extracted using TRIzol (Thermo Fischer Scientific) according to the manufacturer's protocol. As verification of the morphological identification, DNA barcoding was done using standard cytochrome c oxidase subunit 1 (COI) PCR and Sanger sequencing. One microgram of the RNA extract was reverse-transcribed at 42°C for 60 min using M-MLV Reverse Transcriptase (Promega). Universal primers comprising COI_F (5′ GGATTTGGAAATTGATTAGTTCCTT 3′) and COI_R (5′ AAA AATTTTAATTCCAGTTGGAACAGC 3′) were used to amplify a 600-bp PCR product. The PCR reaction mixture contained 0.25 × GC buffer (Kapa Biosystems), 1.5 mM $MgCl_2$, 1 mM dNTPs mix, 0.2 μM of each primer, 1.5 U KAPA Taq DNA polymerase (Kapa Biosystems), and 1 μl of cDNA. The thermal profile of the PCR included 40 cycles of denaturation at 95°C for 30 s, annealing at 50°C for 45 s and elongation at 65°C for 1 min, and a final elongation step at 65°C for 7 min. PCR products were purified using the NucleoSpin Gel and PCR Clean-up purification kit (Macherey-Nagel). Sanger sequencing was done on the PCR product, and the resulting sequence was analysed using the Barcode of Life Data System V4 platform (Ratnasingham and Hebert 2007) and tested against local COI sequences available in the National Center for Biotechnology Information (NCBI) Genbank database (Supplementary Table S2).

2.3 Total RNA next-generation sequencing

Mosquitoes from various collection sites within the Eastern Macedonia and Thrace region were separated into homogenous single-species pools of five well-preserved individuals each, and a total RNA for each pool was extracted using TRIzol (Thermo Fischer Scientific) and a pellet pestle (Eppendorf) according to the manufacturer's protocol. Applying the same methodology as for the

Figure 1. a) Geopolitical map, including in red shade the region of Eastern Macedonia and Thrace, Greece. The area is bordered to the north with Bulgaria and to the East with Turkey. b) Physical map of the collection sites in the Eastern Macedonia and Thrace region of Greece. The collections were organized into five smaller geographical areas Drama (1), Xanthi (2), Rhodopi (3), South Evros (4), and North Evros (5). Location black marks indicate the collection sites.

single-species pools, total RNA was extracted from five heterogenous pools of mixed species comprising 100 mosquito individuals each. The mixed-species pools were assembled using mosquitoes sampled from traps representing three out of five subareas that showed the largest variety of species. These pools represented the populations of the Subareas 2 (Pools A and E), 4 (Pools C and D), and 5 (Pool B) and were additionally enriched with species from other areas in order to increase the diversity for this proof-of-principle study (Fig. 1, Supplementary Table S3). The quality of the RNA preparation was assessed using LabChip GX Touch 24 (PerkinElmer) capillary electrophoresis. Whole transcriptome libraries were prepared from 500 ng of RNA extract using the Ion Total RNA-Seq v2 Core Kit (#4479789, ThermoFisher Scientific) according to the manufacturer's instructions. In brief, the RNA library preparation involved RNA fragmentation, adapter ligation, reverse transcription, and 14 cycles of PCR amplification using an Ion Xpress™ RNA-Seq Barcode 1-16 Kit (#4475485, ThermoFisher Scientific). Quantification of the library was done using a Qubit Fluorometer high-sensitivity kit (ThermoFisher Scientific), and its median size was determined using a LabChip GX Touch 24 (PerkinElmer). The libraries were loaded into an Ion 540 chip (yielding ~80 million single-end reads using the 200-bp chemistry) using the automated Ion Chef System (Thermo Fisher Scientific) and single-end sequencing was carried out on an Ion GeneStudio S5, Ion torrent sequencer (ThermoFisher Scientific). The datasets presented in this study can be found in the Sequence Read Archive SRA database of NCBI (SRA; https://www.ncbi.nlm.nih.gov/sra/—last accessed date 20 September 2021) and are identified by their accession numbers as shown in Supplementary Table S4, whereas the output from the next-generation sequencing (NGS) data analysis is included in Supplementary Table S5.

2.4 Viral genome assembly

Following the NGS procedure, raw sequences of mosquitoes derived from the sampled homogenous (single-species) and heterogenous (mixed-species) pools were used as input for a *de novo* assembly using Trinity (v2.8.5.; Grabherr et al. 2011). The Trinity assembler, which is based on the de Bruijn graph algorithm, produces contigs (set of overlapping DNA segments) that represent alternate transcripts of genes, treating sequences with structural changes (mutations and indels) as isoforms of the same gene. The whole process is performed via three distinct modules, namely Inchworm, Chrysalis, and Butterfly, responsible for creating the assemblies of transcripts, clustering them, and optimizing the de Bruijn graphs, respectively. Due to the probabilistic nature of the algorithm, each sample/pool was submitted to five Trinity assembly runs using the default program parameters, thus maximizing

the possibilities of revealing *bona fide* full-length viral sequences. The output data analysis of the Trinity assembly is included in Supplementary Table S6. The generated and assembled contigs of all the Trinity runs were aligned against the non-redundant (nr) protein database via Basic Local Alignment Search Tool × (BLASTx) (Altschul et al. 1990) using taxonomic search restriction on *Viridae* (taxid:10,239) and annotated by their top BLASTx hit. As Trinity is a stochastic assembler (Haas et al. 2013), nucleotide (nt) sequences corresponding to the same top BLASTx hit were fed into the CAP3 tool, using default parameters, in order to combine contigs produced by Trinity into scaffolds, with the aim of maximizing the efficiency of viral genome assembly (Huang and Madan 1999). The reference-guided virus genome assembly was used when the annotated contigs showed a high similarity to the top BLASTx hits. Viral genome assembly was also aided by genomic, proteomic, and structural data sourced from the ViralZone webpage (ViralZone, https://viralzone.expasy.org/—last accessed date 20 September 2021) (Hulo et al. 2011), with respect to the studied virus families and genera. Lastly, alignment algorithms (Burrows–Wheeler Aligner; Li and Durbin 2009), MAFFt (Katoh 2002) and the Integrated Genomics Viewer (Robinson et al. 2011) software were utilized in order to fine-tune and validate the assembled viral sequences before submitting them to the NCBI GenBank database. The custom bioinformatics pipeline developed for the virus genome assembly is presented in Supplementary Figure S1. All assembled viral sequences of this study can be accessed online using their accession numbers (MW520364—MW520429) as shown in Supplementary Table S7. The relative abundance of each virus segment was calculated using the RNA-Seq by Expectation-Maximization (RSEM) method (Li and Dewey 2011). The relative abundance of segmented viruses was represented by the sum of their individual segments. Index hopping cut-off was set at 1 virus read per 1,000,000 reads. With Ion torrent technology, index hopping has been calculated to reach 0.156 per cent (Palmer et al. 2018).

2.5 Phylogenetic analysis of assembled viruses

All custom scripts and files utilized for virome analysis were deposited to Github (Github, https://github.com/konskons11/MOSQ—last accessed date 6 April 2022). For the phylogenetic analysis of the assembled viruses, the amino acid (aa) sequence of the RNA-dependent RNA polymerase or RNA-dependent RNA polymerase (RdRp; also characterized as segment L or L protein) was mainly used for the construction of the phylogenetic trees. In those cases where the RdRp segment was not assembled, the nucleocapsid (segment S or N) or glycoprotein (segment M or G) aa sequence was used. The length of the input sequence varied depending on the length of the assembled contigs. The respective virus aa sequence, obtained from the ExPASy Translate tool (Gasteiger et al. 2003) according to the standard genetic code, was input to Basic Local Alignment Search Tool p (BLASTp) (Altschul et al. 1990) against the nr protein sequence database of NCBI. BLASTp hits with not less than 20 per cent coverage, and 20 per cent identity were selected in order to download their corresponding and complete protein sequences in FASTA format. Host and geographic origin data of these protein sequences were extracted with the NCBI tool *Entrez-direct* (Kans 2021) using a custom bash-shell script (Github, https://github.com/konskons11/MOSQ, file: NCBI-search.sh—last accessed date 6 April 2022), and the retrieved information was added after each protein sequence name within the FASTA file. All downloaded viral protein sequences, together with our assembled viral protein sequence in FASTA format, were used as input to the NGPhylogeny.fr online tool (NGPhylogeny.fr, https://ngphylogeny.fr/—last accessed date 6 April 2022; Lemoine et al. 2019) for the elucidation of phylogenetic relationships. This was done by running the advanced FastME/OneClick workflow (as proposed by NGPhylogeny.fr), with successive stages of multiple sequence alignment done using MAFFT (Katoh 2002) and alignment refinement using BMGE (Criscuolo and Gribaldo 2010) to provide a phylogenetic reconstruction via the FastME substitution model (Lefort, Desper, and Gascuel 2015). The phylogenetic reconstruction was based on a balanced minimum evolution for 100 bootstrap cycles. Finally, a graphical rendering of the inferred trees was done using the utility Newick Display (Junier and Zdobnov 2010). Trees were exported to the software tool iTol (Letunic and Bork 2019) and rooted using the midpoint rooting method. To be noted, not all the taxa from the inferred trees are presented in the figures, as we emphasized in the depiction of the phylogenetic relationships closer to the viruses identified in this study. A detailed account of the phylogenetic analysis workflow is presented in Supplementary Figure S2. Virus nomenclature, taxonomy, and new species designation were done according to demarcation criteria set by the International Committee on Taxonomy of Viruses (ICTV) (Walker et al. 2020).

2.6 Data visualization and statistical analysis

Variants of viruses in two or more mosquito species with high aa sequence similarity (>95 per cent) to each other after MAFFt alignment (Supplementary Table S8) were plotted as links between mosquito species in the habitat, and these were visualized as a Chord Diagram using the circlize package in R (Gu et al. 2014), with all necessary input files and corresponding scripts available online (Github, https://github.com/konskons11/MOSQ, files: chord.csv, chord2.txt, circlize.R—last accessed date 6 April 2022). Mosquito-related viruses that emerged using BLASTp for the phylogenetic analysis of the 13 identified viral families (*Solemoviridae, Luteoviridae, Peribunyaviridae, Phasmaviridae, Partitiviridae, Iflavir idae, Chrysoviridae, Totiviridae, Orthomyxoviridae, Rhabdoviridae, Virgaviridae, Narnaviridae,* and *Tymoviridae*), and one marked as unclassified, were plotted against host/carrier mosquito genera using a Sankey diagram (Observable, https://observablehq.com/@mbostock/flow-o-matic—last accessed date 20 September 2021).

The Sankey diagram was generated according to the studied input data, which can be accessed online (Github, https://github.com/konskons11/MOSQ, file: sankey.csv—last accessed date 6 April 2022). The dataset was generated after NCBI PubMed search of all combinations between the virus families reported in this study and the genera found in the area of Thrace. Only papers reporting the exact host were used (mixed species samples were excluded). The dataset was enriched after NCBI GenBank search with the same criteria for the ORGANISM and HOST parameters.

Using the number of individuals per species that were included in the mixed-species pools (Supplementary Table S3) and the total number of NGS reads (reads/million) of the respective viruses (single-species virome), we employed a linear regression model to ascertain if there was a predictive value in the host–virome relationship and determine quantitatively the stability of core virome for any given mosquito species. Species represented in more than three mixed-species pools were used for the final quantitative analysis. All custom scripts and files utilized for the linear regression model can be accessed online (Github, https://github.com/konskons11/MOSQ, files: linear_reg-parameters.txt, linear_reg.tsv, linear_reg.R—last accessed date 6 April 2022).

3. Results

In order to analyse the core virome of all the mosquito species in the local ecosystem, mosquitoes were collected during April–October 2018 and April–October 2019 using CDC traps in rural areas of the prefecture of Eastern Macedonia and Thrace, northern Greece (14,157 km^2), which has a wide variety of landscape features, including sea shores, mountains (~2200 m), lakes, lagoons, swamps, rivers, and two large river deltas (Fig. 1). The prefecture shares borders with Bulgaria and Turkey and, in the past, has been significantly affected by vector-borne diseases (Kampen et al. 2003; Nomikou et al. 2009; Papa et al. 2010; Groen et al. 2017). Morphological characterization assisted by COI barcoding resulted in the identification of 24 mosquito species belonging to six genera (Supplementary Table S9). Single-species pools of five individuals each, obtained from different sites within the local ecosystem, were used for total RNA extraction and total RNA NGS (total RNA-Seq). Pools of five individuals have been shown in the past to yield all the required data necessary for core virome analysis (C. Shi et al. 2019). The same procedure was followed for five mixed-species pools of 100 mosquito individuals that were collected from different places within the same ecosystem (Supplementary Table S3).

All field-collected mosquito samples underwent a total RNA-Seq protocol based on the Ion Torrent sequencing platform, followed by five Trinity assembly runs of the obtained NGS data. The results from the sequencing runs and the Trinity assemblies are summarized in Supplementary Tables S5 and S6. Using a custom bioinformatics pipeline (Supplementary Figure S1), we were able to identify 34 viruses, including 15 novel virus species exhibiting only limited phylogenetic relationship to previously known species. All 24 mosquito species showcased a wide range of viruses, which corresponded to the following 13 viral families: *Solemoviridae, Luteoviridae, Peribunyaviridae, Phasmaviridae, Partitiviridae, Iflaviridae, Chrysoviridae, Totiviridae, Orthomyxoviridae, Rhabdoviridae, Virgaviridae, Narnaviridae*, and *Tymoviridae*, and one marked as unclassified at the family level (Fig. 2, Supplementary Table S10). The viruses that were identified and assembled from the single-species pools, together with the five pools of mixed species, were used to build a reference database of core mosquito viromes in the local ecosystem. The NGS reads from all of the single-species pools, as well as the five mixed-species pools, were aligned against the reference database for the identification of viruses with low abundance or viruses that were missed during the BLASTx annotation process. The final table of virus abundance in single-species pools (collared boxes in Fig. 2) had a 92 per cent overlap with the initial virus identification after BLASTx (black dots in Fig. 2). The situation was completely different in the mixed-species pools as the use of the local ecosystem virus reference database revealed the presence of several low-abundance viral signatures. Total virome analysis of the examined mosquito genera and mixed-species mosquito populations are described thoroughly in the following paragraphs. Phylogenetic analysis for each virus in every mosquito species was used to identify similarities with previously known virus species and designation of viruses as novel. High divergence of some non-RdRp proteins, used for phylogenetic tree construction, resulted in low bootstrap values.

3.1 Aedes

The genus *Aedes* was the most species-rich genus examined in this study and included the following eight species: *Ae. albopictus, Ae. detritus, Ae. pulcritarsis, Ae. caspius, Ae. sticticus, Ae. vexans, Ae. Rusticus*, and *Ae. geniculatus*. The last three species did not yield any virus sequences, while all the others yielded virus contigs belonging to the families *Solemoviridae, Luteoviridae, Phasmaviridae, Totiviridae, Orthomyxoviridae, Rhabdoviridae*, and *Narnaviridae* (Fig. 2). Two different viruses, Wenzhou sobemo-like virus 4 strain Nea Chili and Flen bunya-like virus strain Thrace, were detected in *Ae. albopictus* mosquitoes (Figs 2–3, Supplementary Table S10), while *Aedes detritus* carried three different viruses: Atrato sobemo-like virus 1 strain Evros, Whidbey virus strain Thrace, and Ochlerotatus-associated narna-like virus 2 strain Xanthi (Figs 2–4; Supplementary Table S10). An Ochlerotatus-associated narna-like virus 2 strain Xanthi, similar to that found in *Ae. detritus* was found in *Ae. Pulcritarsis* but was not assembled completely due to a low virus abundance (Fig. 2). *Aedes pulcritarsis* mosquitoes carried another member of the sobemo-like viruses, the Atrato sobemo-like virus 4 strain Evros (Figs 2–3, Supplementary Table S10). *Aedes caspius* yielded contigs corresponding to the glycoprotein and the RdRp transcripts of a rather divergent rhabdovirus, the Evros rhabdovirus 1, whereas the rest of its genes (expressing nucleocapsid, matrix protein, and phosphoprotein) were fragmented or missing and thus were not presented in the overall results table (Fig. 2). Although the Evros rhabdovirus 1 glycoprotein and RdRp exhibited high coverage, they showed low aa identity compared to the previously identified Klamath and Ohlsdorf viruses (Fig. 4, Supplementary Table S10). Although we cannot exclude that these two fragments belong to different viruses, the fact that there was no other corresponding sequence in the sample may indicate that they are part of the same genome. The low similarity to other related viruses supports the identification of a novel rhabdovirus. Finally, a member of *Totiviridae* family, the Aedes aegypti totivirus strain Thrace was found in *Ae. sticticus* mosquitoes (Figs 2 and 5, Supplementary Table S10).

3.2 Culex

Mosquitoes of the genus *Culex* comprised five species; *Cx. modestus, Cx. pipiens, Cx. theileri, Cx. impudicus*, and *Cx. perexiguus*. Analysis of species in the Culex genus yielded virus contigs belonging to the families *Solemoviridae, Luteoviridae, Phasmaviridae, Virgaviridae, Totiviridae, Partitiviridae, Orthomyxoviridae, Rhabdoviridae*, and *Chrysoviridae* (Fig. 2). *Culex modestus* yielded contigs belonging to five viruses, but only three viral genomes could be assembled fully, namely the Culex inatomii totivirus strain Thrace, the Culex orthomyxo-like virus strain Thrace, and the Alexandroupolis virga-like virus, whereas the Beihai partiti-like virus 2 strain Thrace and the Hubei chryso-like virus 1 strain Thrace were partially assembled (Figs 2–5, Supplementary Table S10). Notably, the novel Alexandroupolis virga-like virus, despite displaying a significant similarity to the Hubei virga-like virus 2, had a lower than 90 per cent total nt identity to any previously reported virus and may be designated as being a new virgavirus (Fig. 5; Supplementary Table S10). *Culex pipiens* carried a member of the family *Orthomyxoviridae*, the Wuhan Mosquito Virus 6 strain Thrace (Figs 2–3, Supplementary Table S10). In the *Cx. theileri* pool, the novel Thassos sobemo-like virus was detected along with genetic traces of the novel Alexandroupolis virga-like virus, first assembled in *Cx modestus* species (Fig. 2). Thassos sobemo-like virus was assembled completely, showing less than 50 per cent aa similarity to previously identified and related viruses (Fig. 3, Supplementary Table S10). Moreover, another partially assembled variant of the Alexandroupolis virga-like virus, closely related to Hubei virga-like virus 2 (Fig. 5, Supplementary Table S10), was detected in the *Cx. theileri* pool. A virus predicted as belonging to the family *Luteoviridae,*

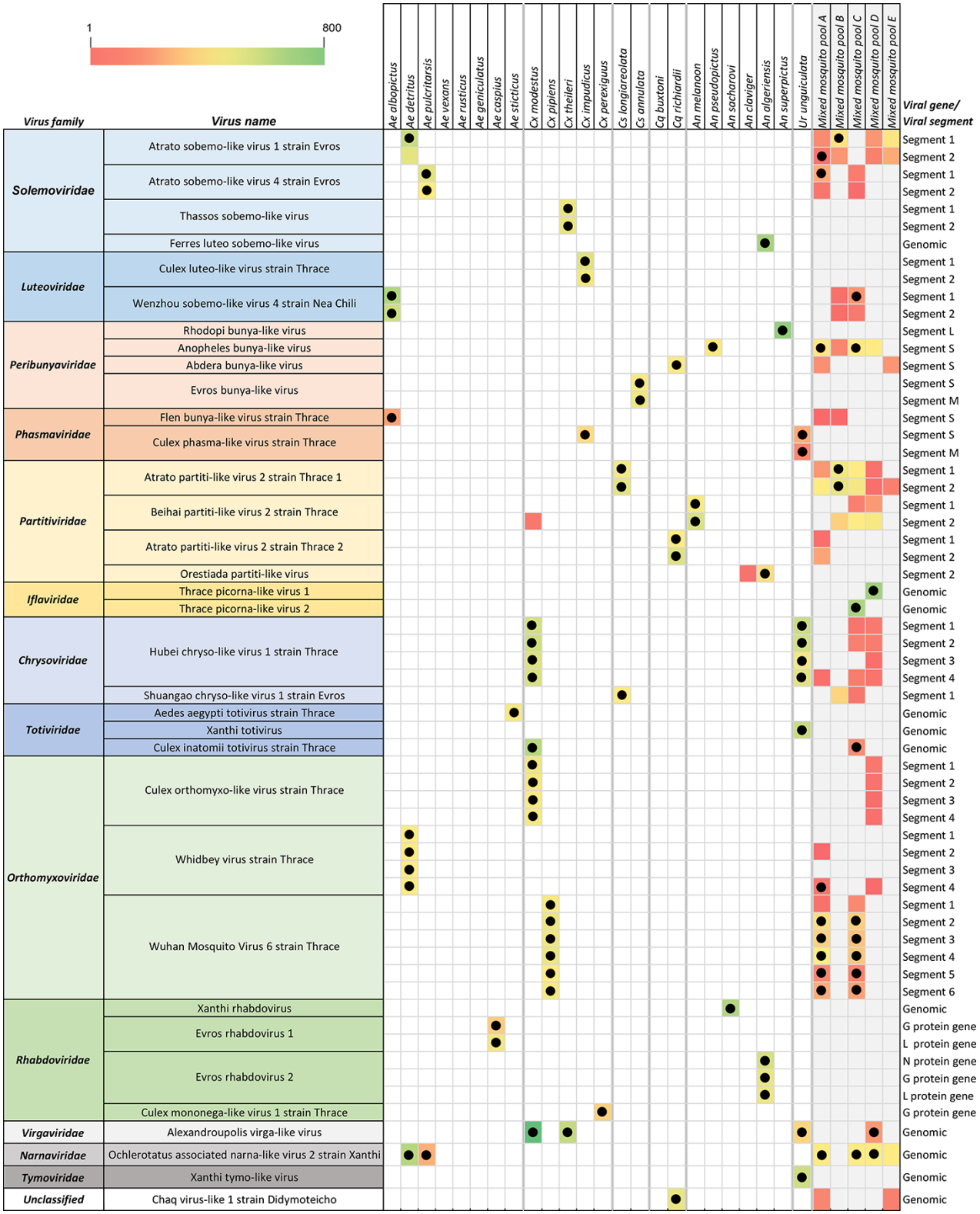

Figure 2. Overall summary of viruses identified in the present study. The viruses are grouped according to the family in which they were predicted to be classified. Viruses identified in the 24 single-species and the 5 mixed-species pools of mosquitoes are reported. Viral genes or segments are noted on the right. Black dots represent the identification of respective virus contigs after Trinity assembly and BLASTx annotation. Coloured square cells (heatmap) depict virus, or virus segment, relative abundance after Burrows–Wheeler Aligner alignment of the respective mosquito total RNA reads on the local virus database generated in the present study. The heatmap corresponds to the viral abundance calculated by the RSEM algorithm and expressed as transcripts per million according to the indicated range.

Culex luteo-like virus strain Thrace (Fig. 2), was detected in the *Cx. impudicus* sample and resembled Culex luteo-like virus (Fig. 3; Supplementary Table S10). Moreover, only the putative nucleocapsid (segment S) of the Culex phasma-like virus strain Thrace was detected and partially assembled from reads in the *Cx. impudicus* pool (Fig. 2) that showed a close similarity to the

Figure 3. *Solemoviridae*, *Luteoviridae*, *Peribunyaviridae*, *Phasmaviridae*, and *Orthomyxoviridae* family phylogenetic trees of the identified viruses in this study distinguished by bold text. Phylogenetic analysis was performed according to the protein molecule indicated between the parentheses after each virus family name. Tree scale is displayed on the upper left corner of each tree. FastME minimum evolution substitution model was utilized as part of the NGPhylogeny.fr methodology. Bootstrap values (blue coloured text) were obtained from 100 bootstrap replicates and only those above 70 are displayed at the start of each node, expressed as percentage (%). Host and country origin information of homologous viruses were also extracted, if applicable, depicted here in purple and black, respectively.

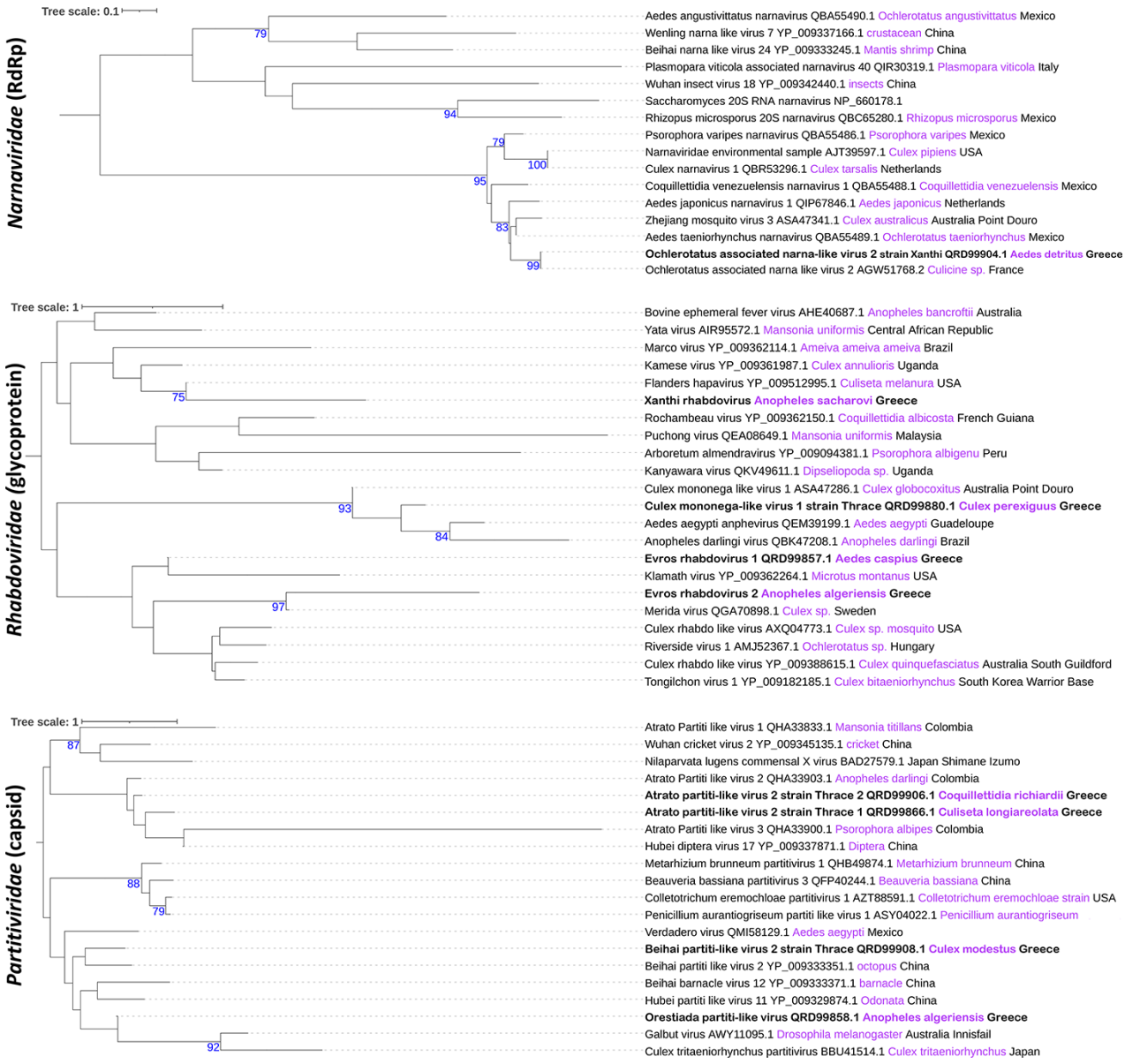

Figure 4. *Narnaviridae, Rhabdoviridae,* and *Partitiviridae* family phylogenetic trees of the identified viruses in this study distinguished by bold text. Tree scale is displayed on the upper left corner of each tree. Phylogenetic analysis was performed according to the protein molecule indicated between the parentheses after each virus family name. FastME minimum evolution substitution model was utilized as part of the NGPhylogeny.fr methodology. Bootstrap values (blue coloured text) were obtained from 100 bootstrap replicates and only those above 70 are displayed at the start of each node, expressed as percentage (%). Host and country origin information of homologous viruses were also extracted, if applicable, depicted here in purple and black respectively.

corresponding nucleocapsid of Culex phasma-like virus (Fig. 3, Supplementary Table S10). Lastly, *Cx. perexiguus* carried the Culex mononega-like virus 1 strain Thrace (Fig. 2) from which only the glycoprotein portion could be assembled successfully (Supplementary Table S10).

3.3 Culiseta

Two species belonging to the genus *Culiseta* were identified: *Cs. longiareolata* and *Cs. annulata*. The species *Cs. longiareolata* carried 2 viruses belonging to the *Partitiviridae* and *Chrysoviridae* families, namely the Atrato partiti-like virus 2 strain Thrace 1 and the Shuangao chryso-like virus 1 strain Evros (Figs 2, 4, and 5; Supplementary Table S10). A novel bunyavirus, the Evros bunya-like virus, was detected in the single-species pool of *Cs. annulata* (Fig. 2), which deviated from known viral sequences. The nucleocapsid (segment S) of the Evros bunya-like virus aligned to an environmental sample metagenome contig of a virus belonging to the family *Peribunyaviridae*, although displaying low aa identity. The glycoprotein (segment M) of Evros bunya-like virus showed similarity to the Culex pseudovishnui bunya-like virus, with 34.99 per cent aa identity (Fig. 3, Supplementary Table S10). Notably, RdRp (segment L) of the novel Evros bunya-like virus could not be assembled, possibly due to low abundance in the samples.

3.4 Coquillettidia

Two mosquito species belonging to the genus *Coquillettidia* were collected, *Cq. buxtoni* and *Cq. richiardii*. *Coquillettidia buxtoni* yielded

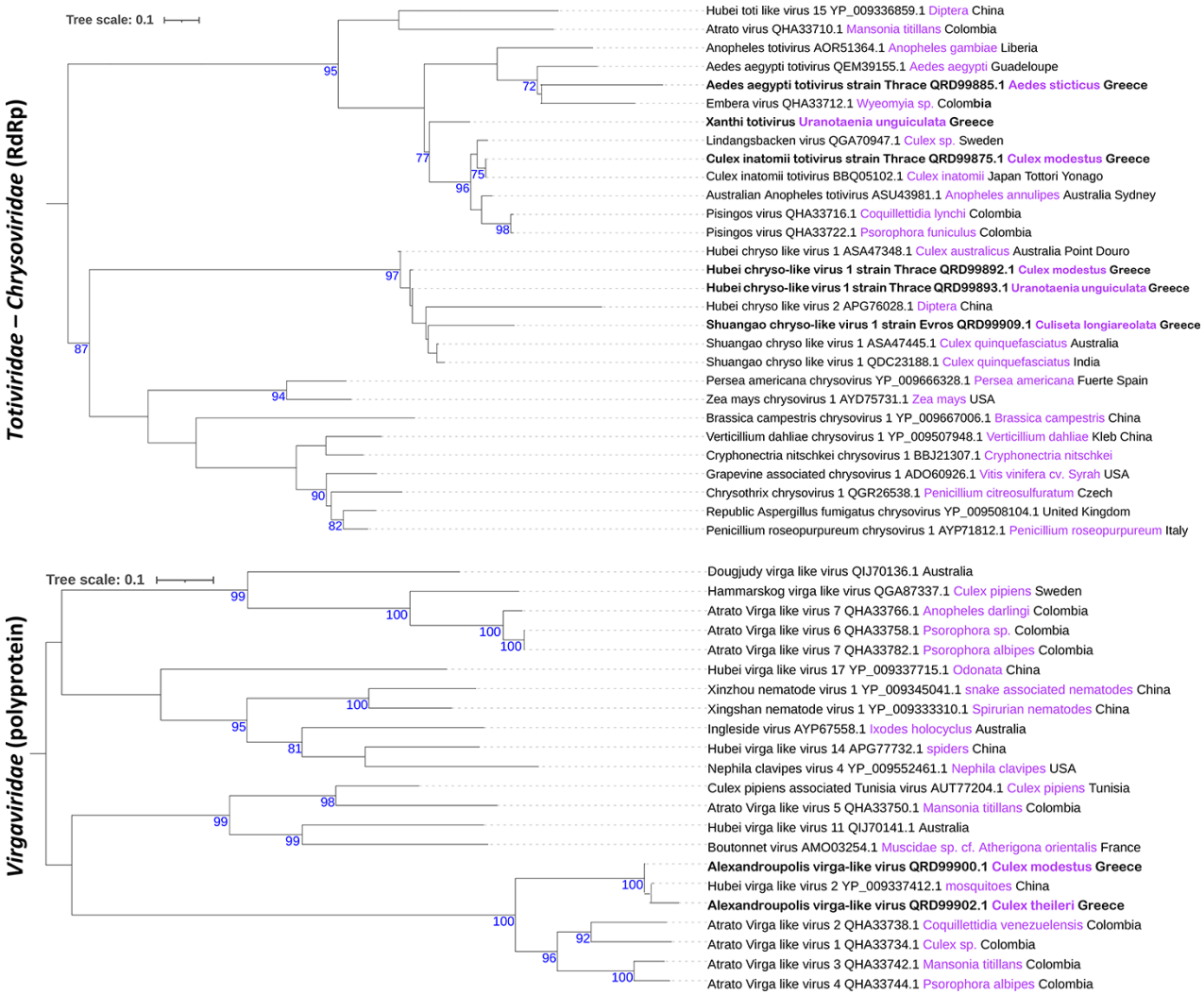

Figure 5. *Totiviridae*, *Chrysoviridae*, and *Virgaviridae* family phylogenetic trees of the identified viruses in this study distinguished by bold text. Tree scale is displayed on the upper left corner of each tree. Phylogenetic analysis was performed according to the protein molecule indicated between the parentheses after each virus family name. FastME minimum evolution substitution model was utilized as part of the NGPhylogeny.fr methodology. Bootstrap values (blue coloured text) were obtained from 100 bootstrap replicates and only those above 70 are displayed at the start of each node, expressed as percentage (%). Host and country origin information of homologous viruses were also extracted, if applicable, depicted here in purple and black, respectively.

no virus contigs, whereas *Cq. richiardii* harboured viruses of the families *Peribunyaviridae* and *Partitiviridae* and an unclassified virus (Fig. 2). The nucleocapsid (segment S) of the novel Abdera bunya-like virus was detected and assembled (Fig. 2). The assembled genome lacked RdRp and glycoprotein segments, possibly due to low abundance in the samples. The nucleocapsid of the Abdera bunya-like virus exhibited only 36.09 per cent aa identity against the corresponding nucleocapsid of Culex pseudovishnui bunya-like virus (Fig. 3; Supplementary Table S10). The Atrato partiti-like virus 2 strain Thrace 2 was retrieved from *Cq. richiardii* mosquitoes and assembled successfully, possessing separate RdRp and capsid protein-encoding genomic segments (Fig. 2) similar to viruses identified before in mosquitoes of the genus *Anopheles*. (Fig. 4; Supplementary Table S10). A strain of a previously identified unclassified virus, the Chaq virus-like 1 strain Didymoteicho, was found within *Cq. richiardii* mosquitoes (Figs. 2 and 6; Supplementary Table S10).

3.5 Anopheles

The genus *Anopheles* was a rich group of mosquitoes and consisted of six species, *An. melanoon*, *An. pseudopictus*, *An. sacharovi*, *An. claviger*, *An. Algeriensis*, and *An. superpictus*. Sequencing of these species yielded virus contigs belonging to the families *Solemoviridae*, *Peribunyaviridae*, *Partitiviridae*, and *Rhabdoviridae* (Fig. 2). *Anopheles melanoon* carried the Beihai partiti-like virus 2 strain Thrace (Fig. 2). Two novel but closely related viruses of the family *Peribunyaviridae* were detected separately in the species *An. pseudopictus* (Anopheles bunya-like virus) and *An. superpictus* (Rhodopi bunya-like virus) (Figs 2–3; Supplementary Table S10). As both viruses presented significantly lower RdRp and nucleocapsid aa identity compared to the species demarcation criteria set by ICTV, they may be designated as novel bunyaviruses. The novel Xanthi rhabdovirus (Fig. 2) was found in *An. sacharovi* mosquitoes and was assembled successfully as a complete genome, possessing five genes, annotated as putative N, M, P, G, and L protein-encoding

genes. (Fig. 4; Supplementary Table S10). As the aa identity to previously known viruses was less than 80 per cent, the Xanthi rhabdovirus may be considered as a new rhabdovirus. *Anopheles algeriensis* carried three novel viruses (Fig. 2). The Ferres luteo sobemo-like virus displayed almost perfect coverage but low levels of aa identity to the Sanxia sobemo-like virus 3 (Fig. 3; Supplementary Table S10). Another virus within the *An. algeriensis* mosquito pool was Evros rhabdovirus 2 (Fig. 2), the genome of which was almost completely assembled, apart from its M and P protein-encoding region. (Fig. 4; Supplementary Table S10). Finally, the Orestiada partiti-like virus was assembled partially (Fig. 2). Only the capsid protein-encoding segment of Orestiada partiti-like virus was assembled successfully and aligned best against the Beihai barnacle virus 12 capsid (Fig. 4; Supplementary Table S10). A similar virus was also detected in the *An. claviger* pool, although it was significantly fragmented.

3.6 Uranotaenia

The species *Ur. unguiculata*, harboured five viruses, belonging to the families *Phasmaviridae*, *Virgaviridae*, *Chrysoviridae*, *Totiviridae*, and *Tymoviridae*. Two of these viruses, which were variants of the Culex phasma-like virus strain Thrace and the novel Alexandroupolis virga-like virus (Fig. 2), had mainly fragmented genomic sequences and therefore could not be assembled fully (Fig. 3, Supplementary Table S10). Another variant of the Hubei chryso-like virus 1 strain Thrace was found within the *Ur. unguiculata* sample pool (Fig. 5; Supplementary Table S10). A novel member of the family *Totiviridae*, the Xanthi Totivirus was also detected and successfully assembled (Fig. 2). The encoded capsid and RdRp proteins of the Xanthi Totivirus resembled the corresponding proteins of the Australian Anopheles totivirus and Culex inatomii totivirus, showing <50 per cent aa identity, which is the threshold for the designation of a new species (Fig. 5; Supplementary Table S10). Lastly, a novel tymovirus, the Xanthi tymo-like virus was retrieved and fully assembled (Fig. 2). Xanthi tymo-like virus RdRp was related to the corresponding Culex pseudovishnui tymo-like virus RdRp with only 54.96 per cent aa identity, far less than the 80 per cent threshold for species demarcation, indicating the novelty of this virus (Fig. 6; Supplementary Table S10).

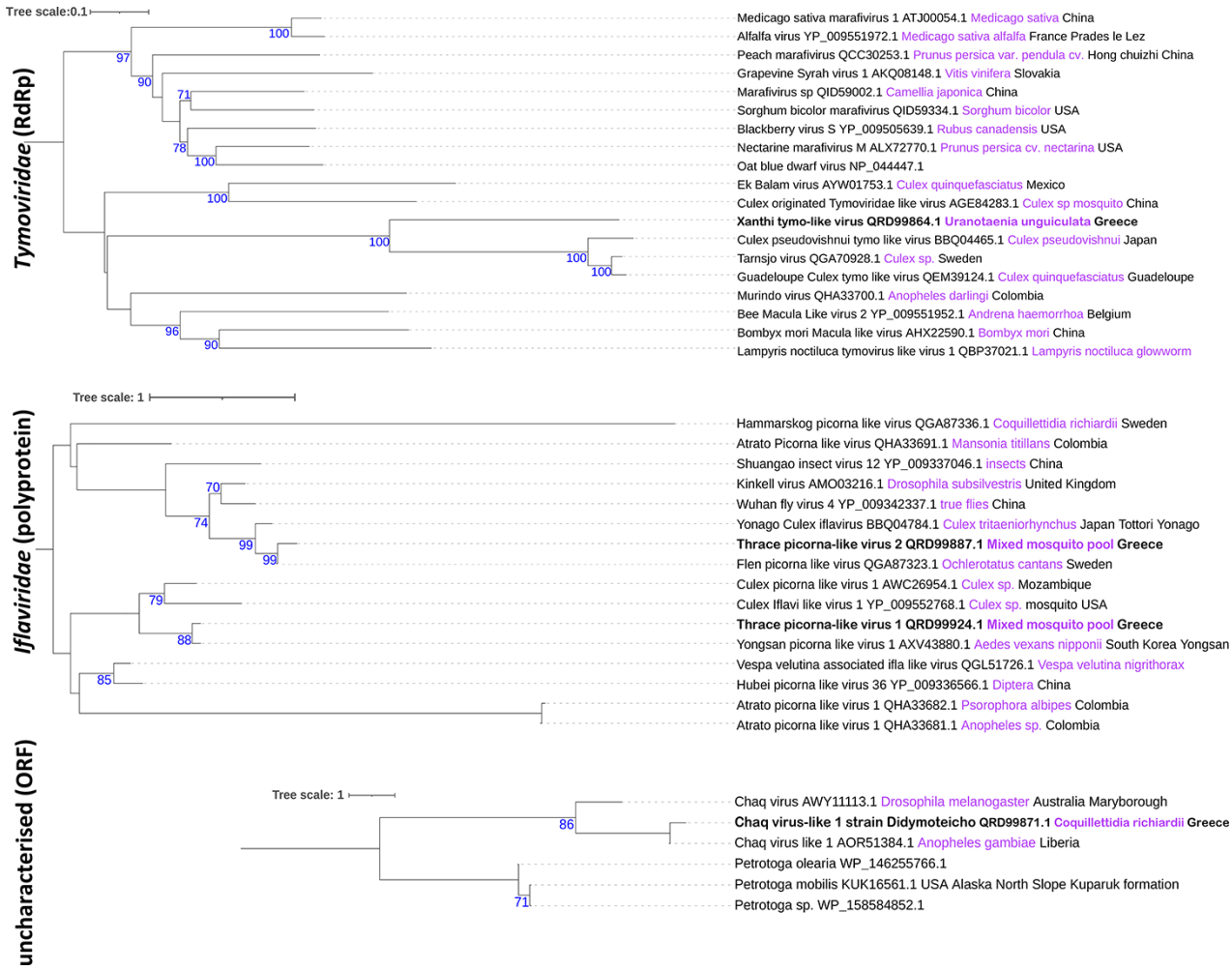

Figure 6. *Tymoviridae*, *Iflaviridae*, and uncharacterized family phylogenetic trees of the identified viruses in this study distinguished by bold text. Tree scale is displayed on the upper left corner of each tree. Phylogenetic analysis was performed according to the protein molecule indicated between the parentheses after each virus family name. FastME minimum evolution substitution model was utilized as part of the NGPhylogeny.fr methodology. Bootstrap values (blue coloured text) were obtained from 100 bootstrap replicates and only those above 70 are displayed at the start of each node, expressed as percentage (%). Host and country origin information of homologous viruses were also extracted, if applicable, depicted here in purple and black, respectively.

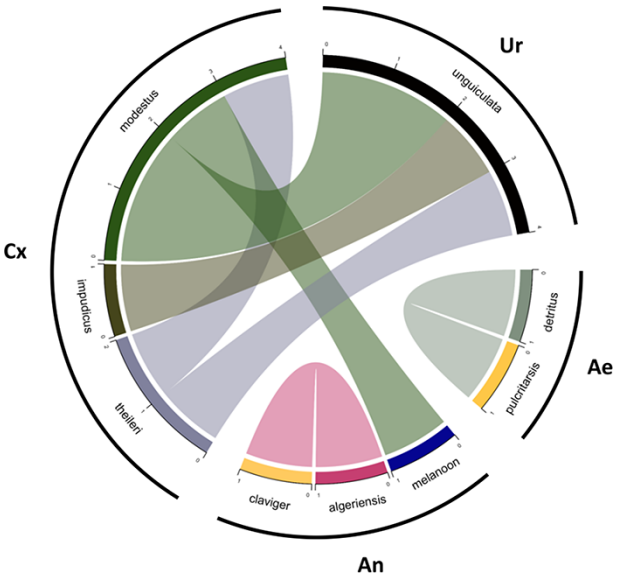

Figure 7. Virus-carrier/host Chord Diagram network in the ecosystem of Eastern Macedonia and Thrace region. Variants of assembled viruses detected in multiple mosquito species, which presented high aa sequence similarity (>95 per cent) to each other, were plotted as links, highlighting the interconnectedness between mosquito species and viruses in the habitat.

3.7 Mixed species mosquito pools

The same procedure, as described above for the single-species pools, was performed on the five mixed-species pools of mosquitoes. The vast majority of BLASTx hits of the contigs assembled from the mixed-species pools corresponded to viruses present in the single-species pools (Fig. 2). In pools C and D, two novel unrelated viruses belonging to the order *Picornavirales*, family *Iflaviridae*, were detected. These novel viruses, the Thrace picorna-like 1 virus and the Thrace picorna-like 2 virus, were assembled at full length due to the high abundance of the respective NGS reads (Fig. 6; Supplementary Table S10). Mapping of the reads of the mixed-species pools on the local virus database revealed a large representation of the individual-species viromes in the virome of the mixed-species pools (Fig. 2). Eighteen out of the 32 viruses (56 per cent) identified in the single-species pools were also represented in the mixed-species pools.

3.8 Host-virus networks in the ecosystem

In order to build an ecosystem-based host-virus network, we compared the core viromes of all mosquito species to identify common paths of virus transmission. Analysis of the core viromes of individual mosquito species showed that certain viruses were shared not only between species of the same genus but also between different genera (Fig. 7). *Aedes detritus* and *Ae. pulcritarsis* shared two variants of the Ochlerotatus-associated narna-like virus 2 strain Xanthi, with 89.74 per cent nt similarity (Supplementary Table S8). *Anopheles algeriensis* and *An. claviger* shared two variants of the Orestiada partiti-like virus segment 2 with 91.50 per cent nt identity (Supplementary Table S8). Rhodopi bunya-like virus and Anopheles bunya-like virus segments found in *An. superpictus* and *An. pseudopictus*, aligned independently to the two segments of a previously identified virus (AXQ04766, AXQ04767) and although they do not share any common segment, were considered to be similar (Fig. 7).

Interestingly, *Cx. modestus* shared common viruses with a variety of species, including *Cx. theileri* (Alexandroupolis virga-like virus, 86.95 per cent nt identity), *An. melanoon* (Beihai partiti-like virus 2 strain Thrace segment 2, 98.57 per cent nt identity), and *Ur. unguiculata* (Hubei chryso-like virus 1 strain Thrace, >87 per cent nt identity per segment, and Alexandroupolis virga-like virus, 90.29 per cent nt identity; Supplementary Table S8). *Uranotaenia unguiculata* emerged also as an important hub of viruses as, apart from the two viruses shared with *Cx. modestus*, it also shared the Alexandroupolis virga-like virus with *Cx. theileri* (78.78 per cent nt identity) and the Culex phasma-like virus strain Thrace with *Cx. impudicus* (91.05 per cent nt identity) (Supplementary Table S8).

3.9 Novel links between viral families and mosquito genera

As representatives of the genera *Culiseta* and *Uranotaenia*, together with a multitude of species within the genera *Aedes*, *Culex*, *Anopheles*, and *Coquillettidia*, were sequenced for the first time, we attempted to identify novel links between virus families and mosquito species. Host-annotated phylogenetic trees of the viruses presented in this study provided data on the associations between mosquito genera and the virus families they carry (Fig. 8). Ten new connections in the global host-virus network at the genus level were identified in this study. This finding highlights the fact that multiple mosquito species may contribute to the circulation of the core virome in any given local ecosystem, although virus genomes may have diverged significantly between mosquito species possibly during adaptation.

3.10 Stability of the core virome of individual species in a local ecosystem

As core virome was proven to be stable in recent studies, it was hypothesized that the composition of the mosquito virome of every species could be stable in a given local ecosystem. Furthermore, it was hypothesized that each species of mosquito virome may be stable in terms not only of quality but also of quantity. Mapping the NGS reads from the mixed-species pools to the local virus database, we quantified the normalized read count for each virus per pool (as performed for individual species). The NGS reads mapping to all virus genomes assigned to a mosquito species (total mosquito virome abundance) was plotted against the number of individuals of this mosquito species in the respective mixed-species pool. Linear regression models applied on the aforementioned plots revealed, in most cases, a statistically significant relationship between the abundance of a mosquito species and the relative abundance of the respective virome within the total virome of the local ecosystem (Fig. 9). Despite the fact that certain species (*Ae. detritus*, *Cq. Richardii*, and *Cx. pipiens*) did not show strong statistical significance, they displayed a similar trend in the linear regression model (Fig. 9). This relationship, which applied to almost all studied mosquito species, may support quantitatively the stability of these mosquito species viromes in the region of Thrace. Further sampling may be required to assess the stability of the mosquito virome of less-abundant mosquito species in the area.

4. Discussion

During the past decade, a great number of mosquito species have been analysed using metagenomics methodologies for the identification of medically important viruses and/within their core

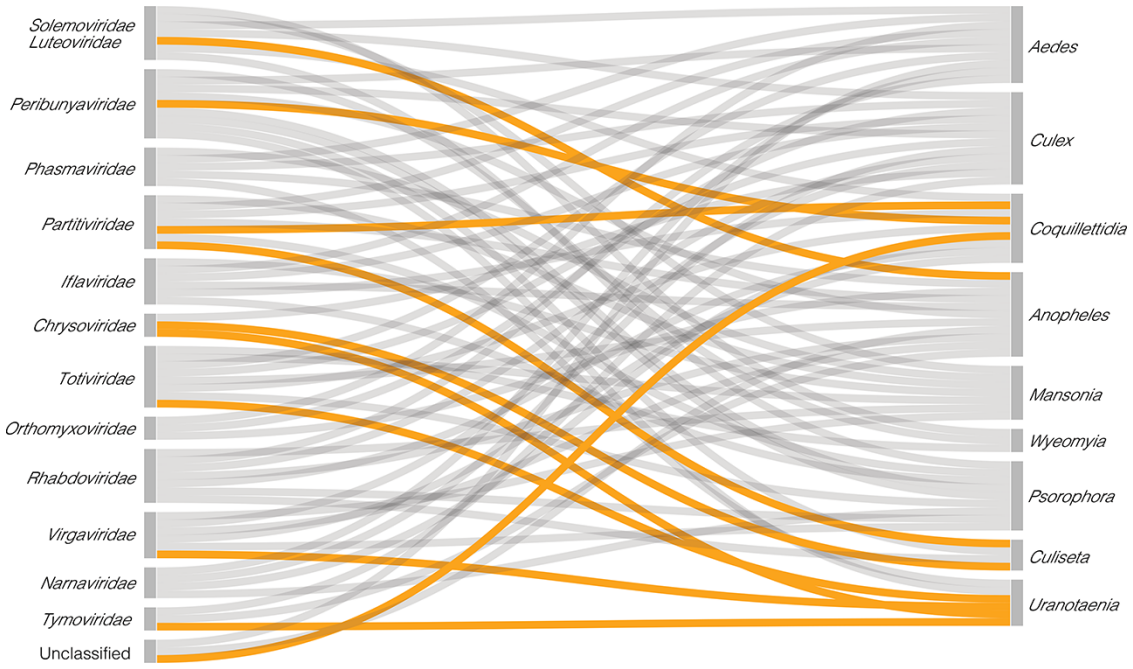

Figure 8. Sankey diagram of the global virus-carrier/host network involving viruses with >20 per cent aa similarity to the viruses identified in the study. On the left side are the virus families and on the right side are the respective carrier/host mosquito genera. Grey chords represent links reported in the literature. Orange chords represent links identified in this study.

virome. Previous studies have focused on a limited number of species, usually one or two, in order to characterize in depth or compare their core viromes in a specific region (Fauver et al. 2016; Colmant et al. 2017; Pettersson et al. 2019; C. Shi et al. 2019; Abbo et al. 2020; Faizah et al. 2020). In our report, we sought to analyse the total mosquito core virome of all mosquito species that occurred in an ecosystem and compare their viromes to seek divergence and convergence points.

Taking into consideration the evolutionary or current crosstalk between species viromes and the fact that multiple mosquito species may play complementary roles in an ecosystem, at either an ecological or public health level, we analysed the total mosquito virome of a local ecosystem. The regions of Eastern Macedonia and Thrace are situated on the southeastern border of Europe with Asia and between the mountain ranges of the Balkan Mountains and the Aegean Sea. Lakes, lagoons, swamps, rivers, and two large river deltas characterize this small geographic region that encompasses a high level of biodiversity (Legakis, Constantinidis, and Petrakis 2018). Indeed, we were able to analyse representatives of 24 different mosquito species, all of which were encountered in varying ratios, in different parts of the local ecosystem.

In-depth analysis of the core viromes of the 24 detected mosquito species revealed the abundant presence of 34 viruses belonging to 13 virus families and one marked as unclassified at the family level. Nineteen viruses were related to previously described isolates or metagenomes. They were identified as sobemo-like viruses, luteo-like viruses, orthomyxo-like viruses, picorna-like viruses, chryso-like viruses, virga-like viruses, and narna-like viruses and displayed more than 85 per cent aa similarity to known viruses (Supplementary Table S10). On the other hand, 15 viruses showed significant divergence when compared to their best match in the NCBI GenBank protein database and were hypothesized to be novel, according to the virus species demarcation criteria set out by ICTV for each virus family. Two novel bunyaviruses (Abdera and Evros bunya-like) were identified in the virome of the mosquito species *Cq. richiardii* and *Cs. annulate*, both of which formed a distinct group within the bunyaviruses tree, although they were located within the branch related to viruses from the mosquito genus *Culex* (Sadeghi et al. 2018; Faizah et al. 2020) (Fig. 3). The identification of a bunyavirus in the mosquito genus *Culiseta*, although never reported as a metagenomics finding, has been reported in the past (McLean et al. 1977). A mosquito species belonging to the genus *Uranotaenia* was analysed for the first time, yielding both a novel totivirus (Xanthi totivirus) that was located in a cluster of mosquito-related totiviruses, previously reported in the mosquito genera *Coquillettidia, Anopheles,* and *Culex* (Colmant et al. 2017; Pettersson et al. 2019; Faizah et al. 2020) (Fig. 5), and a novel tymo-like virus located in a distinct cluster of *Culex*-related viruses (Pettersson et al. 2019; C. Shi et al. 2019; Faizah et al. 2020) (Fig. 6). Another novel totivirus (Aedes aegypti totivirus strain Thrace) was identified in the virome of *Ae. sticticus* that clustered together with other mosquito-related totiviruses found in the mosquito genera *Aedes, Anopheles,* and *Wyeomyia* (Fauver et al. 2016; C. Shi et al. 2019) (Fig. 5). Interestingly, a novel partitivirus (Orestiada partiti-like virus), identified in *An. algeriensis*, formed an outgroup in the phylogenetic tree that was located closer to viruses from other insects (Fig. 4). A partitivirus found in *Cq. richiardii* was present together with a strain of Chaq virus, previously identified as a satellite virus of Galbut partitivirus from drosophila (Cross et al. 2020). However, co-occurrence studies are required to verify this relationship. Finally, three novel rhabdoviruses (Xanthi rhabdovirus, Evros rhabdovirus 1 and 2) were identified in mosquito species of two genera, *Ae. caspius, An. Algeriensis,* and *An. sacharovi* (Fig. 4).

The core viromes identified here were very distinct among mosquito species, although certain viruses, such as the Culex phasma-like virus strain Thrace, Ochlerotatus-associated narna-like virus 2 strain Xanthi, Orestiada partiti-like virus, Beihai partiti-like virus 2 strain Thrace, Alexandroupolis virga-like, and

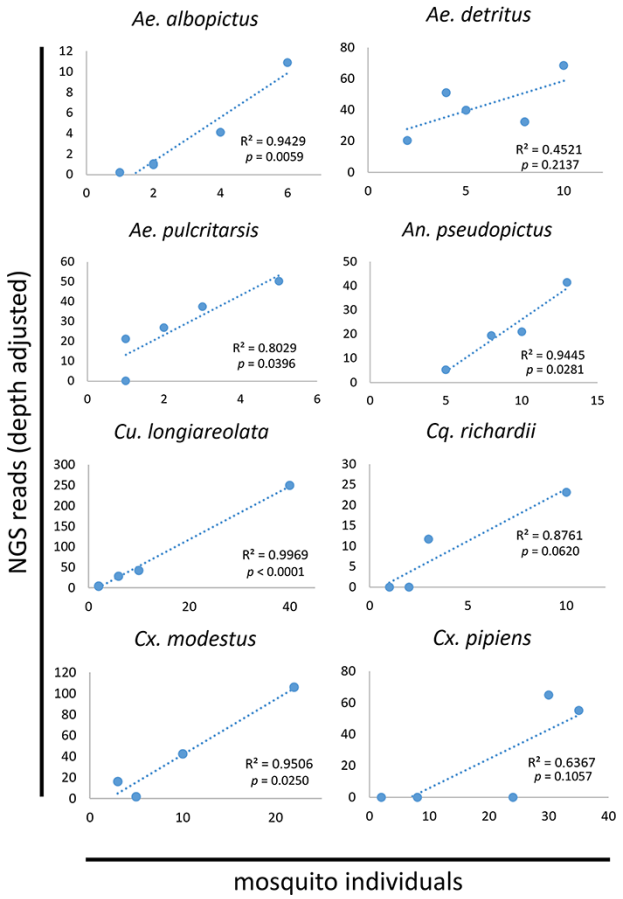

Figure 9. Scatter plots of the number of individuals of different species that were included in mixed-species pools (A–E) versus the total count of NGS reads of the viruses, corresponding to the respective mosquito species, in the same mixed-species pool. Linear regression models were applied with the indicated statistics.

Hubei chryso-like virus 1 strain Thrace, were very similar between mosquito species of the same genus or even between species of different mosquito genera. Despite the fact that these viruses were shared between different mosquito species, they were substantially divergent (~80–90 per cent nt similarity) and are best considered as distinct variants. This difference in similarity may argue against virus sharing between mosquito species in the local ecosystem but may rather indicate that it occurred as a consequence of a recent adaptation event (Grubaugh et al. 2016). As an exception, Beihai partiti-like virus 2 strain Thrace shared between *An. melanoon* and *Cx. modestus* showed almost complete similarity between the different mosquito species and genera, despite the evolutionary distance between these two mosquito species. Core virome genomes are often conserved and stable in either relatively small places, such as Guadeloupe, or large geographical areas such as Australia, Southeast Asia, and Northern Europe (M. Shi et al. 2017; Pettersson et al. 2019; C. Shi et al. 2019). The stability of the core virome between individual members of *Ae. aegypti* and *Cx. quinquefasciatus* occurring in the same ecosystem has been shown recently, where most of the individuals encompassed the same set of viral genomes (C. Shi et al. 2019). Small differences in a mosquito species core virome were recorded in different habitats of Northern Europe (Pettersson et al. 2019). The core virome seems to characterize the individual mosquito species despite overlaps that occur between species of the same genus (M. Shi et al. 2017; Pettersson et al. 2019), whereas there is still an open question as to whether the virome of a species differs significantly between different ecological niches within a local ecosystem. Although we identified links between species, which may signify transmission and adaptation of viruses between different taxa, the vast majority of virus genomes were unique in each mosquito species studied. This might reflect the well-reported existence of restricted tropism of viruses in certain hosts (Webby, Hoffmann, and Webster 2004), but this needs to be further supported by studies involving larger groups of mosquito individuals derived from different regions and different collection years.

Recent reports on the virome of field-collected *Ae. aegypti* and *Cx. quinquefasciatus* in Guadalupe and both laboratory-grown and field-collected *Ae. albopictus* mosquitoes revealed that each mosquito carried a specific subset of viruses that were stable among different individuals (C. Shi et al. 2019, 2020). Furthermore, the stability of the virome was apparent across developmental stages of *Ae. albopictus* (C. Shi et al. 2020) possibly due to vertical transmission. As the core virome seems to be stable in several species, we hypothesized that this remarkable stability may also be reflected quantitively among the mosquito species present in Thrace. In order to quantitatively assess the stability of virome in the most abundant species in the area of Thrace, we evaluated the relationship between the number of individuals of these species, in mixed-species pools of 100 individuals, and the relative abundance of the respective viruses (as identified in the single-species pools; Fig. 9). Despite occasional exchanges of viruses between species, it is apparent that the linear relationship between mosquitoes and their respective viromes further supported the notion of a stable core virome. The virus links between mosquito species are either quantitatively negligible or less apparent during read-mapping as there is still significant nt sequence divergence between the different strains.

5. Conclusions

Our study aimed to assess globally the core mosquito virome in a diverse ecosystem. We identified the presence of 34 viruses, 15 of which were novel, while novel relationships between mosquito species and virus families were also identified. These novel relationships were revealed due to RNA sequencing being used for the first time on a wide range of different species that had not been analysed in the past. Finally, our data provide a quantitative measure of the core virome stability in the most abundant mosquito species of the area, further supporting the idea of a stable core virome in mosquitoes.

Data availability

The raw sequencing datasets for the current study are available in the NCBI Sequence Read Archive repository, under the Bioproject with accession code PRJNA681,030 (NCBI BioProject database, www.ncbi.nlm.nih.gov/bioproject/681030 - last accessed date 20 September 2021). NCBI GenBank accession numbers for the viruses identified in the study can be found on Supplementary Table S7. All custom scripts developed exclusively for the purposes of this study were uploaded to Github and can be accessed online (Github, https://github.com/konskons11/MOSQ - last accessed date 6 April 2022)

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Acknowledgements

We would like to thank Professor Penelope Mavromara for helpful discussions.

Funding

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-5000).

Conflict of interest: The authors declare that they have no competing interests.

Ethics approval

Ethics approval and consent to participate.

Consent for publication

Not applicable.

Author contributions

I.K. designed the study. I.K. and S.V. obtained funding for the project. A.N. conducted fieldwork and collected samples. M.d.C.W. and M.G.R.F. performed morphological species identification. Ko.K., A.K., and Ka.K. performed experiments. K.K. and N.D. performed the analysis. K.K., N.D., and I.K. wrote the manuscript with input from all authors.

References

Abbo, S. R. et al. (2020) 'The Invasive Asian Bush Mosquito Aedes Japonicus Found in the Netherlands Can Experimentally Transmit Zika Virus and Usutu Virus', *PLoS Neglected Tropical Diseases*, 14, e0008217.

Agboli, E. et al. (2019) 'Mosquito-Specific Viruses-Transmission and Interaction', *Viruses*, 11: E873.

Altschul, S. F. et al. (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.

Atoni, E. et al. (2020) 'A Dataset of Distribution and Diversity of Mosquito-associated Viruses and Their Mosquito Vectors in China', *Scientific Data*, 7, 342.

Colmant, A. M. G. et al. (2017) 'Discovery of New Orbiviruses and Totivirus from Anopheles Mosquitoes in Eastern Australia', *Archives of Virology*, 162, 3529–34.

Criscuolo, A., and Gribaldo, S. (2010) 'BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments', *BMC Evolutionary Biology*, 10: 210.

Cross, S. T. et al. (2020) 'Partitiviruses Infecting Drosophila Melanogaster and Aedes Aegypti Exhibit Efficient Biparental Vertical Transmission', *Journal of Virology*, 94, e01070–20.

Dahmana, H., and Mediannikov, O. (2020) 'Mosquito-Borne Diseases Emergence/Resurgence and How to Effectively Control It Biologically', *Pathogens*, 9: 310.

de Brito, T. F. et al. (2021) 'Transovarial Transmission of a Core Virome in the Chagas Disease Vector Rhodnius Prolixus', *PLoS Pathogens*, 17, e1009780.

Elbers, A. R. W., Koenraadt, C. J. M., and Meiswinkel, R. (2015) 'Mosquitoes and Culicoides Biting Midges: Vector Range and the Influence of Climate Change', *Revue Scientifique et Technique (International Office of Epizootics)*, 34: 123–37.

Faizah, A. N. et al. (2020) 'Deciphering the Virome of Culex Vishnui Subgroup Mosquitoes, the Major Vectors of Japanese Encephalitis, in Japan', *Viruses*, 12, E264.

Fauver, J. R. et al. (2016) 'West African Anopheles Gambiae Mosquitoes Harbor a Taxonomically Diverse Virome Including New Insect-specific Flaviviruses, Mononegaviruses, and Totiviruses', *Virology*, 498, 288–99.

Fetters, A. M. et al. (2022) 'The Pollen Virome of Wild Plants and Its Association with Variation in Floral Traits and Land Use', *Nature Communications*, 13, 523.

Gasteiger, E. et al. (2003) 'ExPASy: The Proteomics Server for In-depth Protein Knowledge and Analysis', *Nucleic Acids Research*, 31: 3784–8.

Glick, J. I. (1992). 'Illustrated Key to the Female Anopheles of Southwestern Asia and Egypt (Diptera: Culicidae)', *Mosquito Systematics (USA)*.

Grabherr, M. G. et al. (2011) 'Trinity: Reconstructing a Full-length Transcriptome without a Genome from RNA-Seq Data', *Nature Biotechnology*, 29, 644–52.

Groen, T. A. et al. (2017) 'Ecology of West Nile Virus across Four European Countries: Empirical Modelling of the Culex Pipiens Abundance Dynamics as a Function of Weather', *Parasites & Vectors*, 10, 524.

Grubaugh, N. D. et al. (2016) 'Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching', *Cell Host & Microbe*, 19, 481–92.

Gu, Z. et al. (2014) 'Circlize Implements and Enhances Circular Visualization in R', *Bioinformatics (Oxford, England)*, 30: 2811–2.

Gunay, F., Picard, M., and Robert, V. (2016) 'Interactive Identification Key for Female Mosquitoes (Diptera: Culicidae) of Euro-Mediterranean and Black Sea Regions', *International Journal of Infectious Diseases*, 53: 111.

Haas, B. J. et al. (2013) 'De Novo Transcript Sequence Reconstruction from RNA-seq Using the Trinity Platform for Reference Generation and Analysis', *Nature Protocols*, 8, 1494–512.

Hall-Mendelin, S. et al. (2016) 'The Insect-specific Palm Creek Virus Modulates West Nile Virus Infection in and Transmission by Australian Mosquitoes', *Parasites & Vectors*, 9: 414.

Harbach, R. E., ed. (2018) *Culicipedia: Species-group, Genus-group and Family-group Names in Culicidae (Diptera)*. Wallingford: CABI.

Haynes, M., and Rohwer, F. (2011) 'The Human Virome', In: Nelson, K. E. (ed.) *Metagenomics of the Human Body*, pp. 63–77. Springer: New York, NY.

Hobson-Peters, J. et al. (2013) 'A New Insect-specific Flavivirus from Northern Australia Suppresses Replication of West Nile Virus and Murray Valley Encephalitis Virus in Co-infected Mosquito Cells', *PLoS One*, 8, e56534.

Huang, X., and Madan, A. (1999) 'CAP3: A DNA Sequence Assembly Program', *Genome Research*, 9: 868–77.

Hulo, C. et al. (2011) 'ViralZone: A Knowledge Resource to Understand Virus Diversity', *Nucleic Acids Research*, 39: D576–82.

Jia, J. et al. (2021) 'Interannual Dynamics, Diversity and Evolution of the Virome in Sclerotinia Sclerotiorum from a Single Crop Field', *Virus Evolution*, 7, veab032.

Junier, T., and Zdobnov, E. M. (2010) 'The Newick Utilities: High-throughput Phylogenetic Tree Processing in the Unix Shell', *Bioinformatics*, 26: 1669–70.

Kampen, H. et al. (2003) 'Individual Cases of Autochthonous Malaria in Evros Province, Northern Greece: Entomological Aspects', *Parasitology Research*, 89: 252–8.

Kans, J. (2021) *Entrez Direct: E-utilities on the Unix Command Line. Entrez Programming Utilities Help [Internet]*. National Center for Biotechnology Information (US): Bethesda (MD).

Katoh, K. (2002) 'MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform', *Nucleic Acids Research*, 30: 3059–66.

Lefort, V., Desper, R., and Gascuel, O. (2015) 'FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program', *Molecular Biology and Evolution*, 32: 2798.

Legakis, A., Constantinidis, T., and Petrakis, P. V. (2018) 'Biodiversity in Greece: Selected Countries in Europe'. In: Global Biodiversity. 1st ed., pp. 71–113. Apple Academic Press: Florida, USA.

Lemoine, F. et al. (2019) 'NGPhylogeny.fr: New Generation Phylogenetic Services for Non-specialists', *Nucleic Acids Research*, 47: W260–5.

Letunic, I., and Bork, P. (2019) 'Interactive Tree of Life (Itol) V4: Recent Updates and New Developments', *Nucleic Acids Research*, 47: W256–9.

Li, B., and Dewey, C. N. (2011) 'RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome', *BMC Bioinformatics*, 12: 323.

Li, C.-X. et al. (2015) 'Unprecedented Genomic Diversity of RNA Viruses in Arthropods Reveals the Ancestry of Negative-sense RNA Viruses', *eLife*, 4: e05378.

Li, H., and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform', *Bioinformatics*, 25: 1754–60.

Mayer, S. V., Tesh, R. B., and Vasilakis, N. (2017) 'The Emergence of Arthropod-borne Viral Diseases: A Global Prospective on Dengue, Chikungunya and Zika Fevers', *Acta Tropica*, 166: 155–63.

McLean, D. M. et al. (1977) 'Bunyavirus Isolations from Mosquitoes in the Western Canadian Arctic', *The Journal of Hygiene*, 79: 61–71.

Neu, A. T., Allen, E. E., and Roy, K. (2021) 'Defining and Quantifying the Core Microbiome: Challenges and Prospects', *Proceedings of the National Academy of Sciences*, 118: e2104429118.

Nomikou, K. et al. (2009) 'Evolution and Phylogenetic Analysis of Full-Length VP3 Genes of Eastern Mediterranean Bluetongue Virus Isolates', *PLoS One*, 4, e6437.

Öhlund, P., Lundén, H., and Blomström, A.-L. (2019) 'Insect-specific Virus Evolution and Potential Effects on Vector Competence', *Virus Genes*, 55: 127–37.

Palmer, J. M. et al. (2018) 'Non-biological Synthetic Spike-in Controls and the AMPtk Software Pipeline Improve Mycobiome Data', *PeerJ*, 6: e4925.

Papa, A. et al. (2010) 'Emergence of Crimean-Congo Haemorrhagic Fever in Greece', *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 16: 843–7.

Pettersson, J. H.-O. et al. (2019) 'Meta-Transcriptomic Comparison of the RNA Viromes of the Mosquito Vectors Culex Pipiens and Culex Torrentium in Northern Europe', *Viruses*, 11: E1033.

Pyke, A. T. et al. (2021) 'Uncovering the Genetic Diversity within the Aedes Notoscriptus Virome and Isolation of New Viruses from This Highly Urbanised and Invasive Mosquito', *Virus Evolution*, 7: veab082.

Ratnasingham, S., and Hebert, P. (2007) 'BOLD: The Barcode of Life Data System (www.barcodinglife.org)', *Molecular Ecology Notes*, 7: 355–64.

Robinson, J. T. et al. (2011) 'Integrative Genomics Viewer', *Nature Biotechnology*, 29: 24–6.

Rocklöv, J., and Dubrow, R. (2020) 'Climate Change: An Enduring Challenge for Vector-borne Disease Prevention and Control', *Nature Immunology*, 21: 479–83.

Roundy, C. M. et al. (2017) 'Insect-Specific Viruses: A Historical Overview and Recent Developments', *Advances in Virus Research*, 98: 119–46.

Sadeghi, M. et al. (2018) 'Virome of >12 Thousand Culex Mosquitoes from Throughout California', *Virology*, 523: 74–88.

Samanidou-Voyadjoglou, A., and Harbach, R. (2011) 'Keys to the adult female mosquitoes (Culicidae) of Greece', *European Mosquito Bulletin*, 10: 13–20.

Samanidou-Voyadjoglou, A. (2001) 'The mosquitoes of Greece'. *In: Morphology, Biology, Public Health, Determination keys, Control*. Agrotypos Publications: Athens, Greece.

Shi, C. et al. (2019) 'Stable Distinct Core Eukaryotic Viromes in Different Mosquito Species from Guadeloupe, Using Single Mosquito Viral Metagenomics', *Microbiome*, 7, 121.

Shi, C. et al. (2020) 'Stability of the Virome in Lab- and Field-Collected Aedes Albopictus Mosquitoes across Different Developmental Stages and Possible Core Viruses in the Publicly Available Virome Data of Aedes Mosquitoes', *mSystems*, 5, e00640–20.

Shi, M. et al. (2016) 'Redefining the Invertebrate RNA Virosphere', *Nature*, 540, 539–43.

Shi, M. et al. (2017) 'High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia', *Journal of Virology*, 91: e00680–17.

Vasilakis, N., and Tesh, R. B. (2015) 'Insect-specific Viruses and Their Potential Impact on Arbovirus Transmission', *Current Opinion in Virology*, 15: 69–74.

Viglietta, M. et al. (2021) 'Vector Specificity of Arbovirus Transmission', *Frontiers in Microbiology*, 12: 773211.

Walker, P. J. et al. (2020) 'Changes to Virus Taxonomy and the Statutes Ratified by the International Committee on Taxonomy of Viruses (2020)', *Archives of Virology*, 165, 2737–48.

Webby, R., Hoffmann, E., and Webster, R. (2004) 'Molecular Constraints to Interspecies Transmission of Viral Pathogens', *Nature Medicine*, 10: S77–81.

Wilkerson, R. C. et al. (2015) 'Making Mosquito Taxonomy Useful: A Stable Classification of Tribe Aedini that Balances Utility with Current Knowledge of Evolutionary Relationships', *PLoS One*, 10: e0133602.