bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The virome of *Drosophila suzukii*, an invasive pest of soft fruit

Nathan C. Medd[1], Simon Fellous[2], Madoka Nakai[3], Anne Xuéreb[2], Jerry V. Cross[4] and Darren J. Obbard[1,5]

[1]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

[2]INRA, Montpellier, France

[3]Tokyo University of Horticulture and Technology, Fuchu, Tokyo

[4]NIAB-EMR, East Malling, Kent, UK

[5]Centre for immunity infection and evolution, University of Edinburgh, Edinburgh, UK

Abstract

Drosophila suzukii (Matsumura) is an invasive dipteran pest of soft fruits that causes substantial damage to the horticultural crops of Europe and North America. As conventional control of this pest is challenging, an environmentally benign microbial biopesticide is highly desirable. A lethal virus could prove to be such a control agent. Here we use a metatranscriptomic approach to discover new viruses infecting this fly and its larvae, in both its native (Japanese) and invasive (British and French) ranges. We describe 18 new RNA viruses, including members of the Picornavirales, Mononegavirales, Bunyavirales, Chuviruses, Nodaviridae, Tombusviridae, Reoviridae, and Nidovirales, and discuss their phylogenetic relationships with previously known viruses. We also detect 18 previously described viruses of other *Drosophila* species that show potential association with *D. suzukii* in the wild.

Importance

Drosophila suzukii is one of the most damaging and costly pests to invade temperate horticultural regions in recent history. Current control methods rely heavily on labour intensive cultural control and environmentally harmful chemical controls. A thorough exploration of the pathogens infecting this pest is the first step on the road to the development of an effective biopesticide. We hope these data will facilitate the onward development of such control agents. In addition, because *D. suzukii* is closely related to the well-studied model species *D. melanogaster*, our work provides a valuable comparative dataset for the study of viruses in the model family Drosophilidae.

Introduction

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Drosophila suzukii (Matsamura) is an invasive dipteran pest of soft fruit belonging to the paraphyletic subgenus Sophophora. Unusual to the genus, the larvae of this species are well adapted to feeding on ripe fruit still on the tree. Adult females possess a heavily sclerotized, saw-like ovipositor that allows oviposition under the skin of ripening fruit (Atallah et al., 2014) and their olfactory system is adapted to respond to fruit rather than microbe volatiles (Karageorgi et al., 2017). These evolutionary innovations may aid the establishment of this species in novel habitats across the globe. First described in Japan in 1916 (Kanzawa, 1935, Matsumura, 1931), *D. suzukii* was reported to be widely distributed in Japan shortly after (Kanzawa, 1939). It was recorded across Asia during the last century (Peng, 1937, Kang and Moon, 1968, Parshad and Duggal, 1965, Okada, 1976, Toda, 1991, Sidorenko, 1992, Amin ud Din et al., 2005), with the first records outside of Asia coming from Hawaii in the 1980's (Kaneshiro, 1983). Since its detection in 2008 in the southern states of the USA (Bolda, 2008) and Spain (Calabria et al., 2012), *D. suzukii* has spread northwards, and was recorded for the first time in the UK in 2012 (Harris and Shaw, 2014). Records now stretch from Sweden (Manduric, 2017) to Argentina (Lue et al., 2017), with secondary invasions thought to be responsible for populations detected in South America and the Indian Ocean Islands (Fraimout et al., 2017).

The damage *D. suzukii* has caused in the fruit growing regions of these countries has driven interest in many aspects of the pest's biology, primarily to improve control methods (Asplen et al., 2015). Conventional chemical control of *D. suzukii* is challenging becasue oviposition occurs so close to harvest that management of pesticide residues during crop treatment is of concern (Swoboda-Bhattarai and Burrack, 2015). *D. suzukii* also has a broad host range allowing it to exploit natural refugia (Walsh et al., 2011). An effective biological control agent of *D. suzukii*, compatible with integrated management techniques (Stern et al., 1959), would therefore be highly desirable to horticulturalists worldwide.

Entomopathogenic viruses have the potential for use as environmentally benign, species-specific biological control agents, with certain groups of viruses being used to effectively control insect pests in a range of settings (Hunter-Fujita et al., 1998). The most successful viral control agents to-date are members of the Baculoviridae, with the nuclear polyhedrosis viruses (NPVs) and granulosis viruses (GVs) finding commercial success against lepidopteran pests in forestry and orchard crops, respectively. These viruses produce polyhedrin occlusion bodies that encase infectious virions during the dispersal stage of the viruses' lifecycle. These protein occlusions protect the virus from environmental degradation, and prolong infectivity in the environment (Bishop et al., 1988, Elgee, 1975). For this reason these viruses have been the focus of viral biopesticide development since their first commercial use in 1975 (Shieh and Bohmfalk, 1980). However, despite the relative success of the Baculoviridae, other viral taxa have also been advocated for control purposes. For example, members of the Nudiviridae (Huger, 2005), Reoviridae (Peng et al., 2000), and Parvoviridae (Bergoin and Tijssen, 1998) have shown promise as control agents, despite not achieving commercial success.

As well as identifying some of the natural enemies that could be harnessed to control *D. suzukii* populations, understanding the nature of viral infections in this species may help us better understand the reasons for its geographical spread and survival (Mitchell and Power, 2003, Torchin et al., 2003, Colautti et al., 2004). In particular, the changing pathogen environment that

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

invasive species encounter in the process of invasion is of interest to the field of ecological immunology, in that theories predicting the relative investment an organism should make in its immunity will be based upon the diversity of pathogens experienced its new range (Joshi and Vrieling, 2005, Colautti et al., 2004). Thorough surveys of pathogen diversity in wild invaders, however, remain relatively rare (but see Liu and Stirling, 2006). The genus *Drosophila* is one of the few invertebrate genera in which wild viral pathogen diversity has been explored, with recent comprehensive virus discovery studies in the genus describing over 50 new viruses (Webster et al., 2016, Webster et al., 2015). Furthermore, a history of intensive investigation into the antiviral immunity of *D. melanogaster* (Xu and Cherry, 2014, Sabin et al., 2010, Huszar and Imler, 2008, Kemp and Imler, 2009, Mussabekova et al., 2017, Bronkhorst and van Rij, 2014, Zambon et al., 2006), means that the viruses of *D. suzukii* may prove to be a valuable comparative system for the study of immune system evolution.

Here we report the results of a metatransciptomic survey of virus-like sequences associated with *D. suzukii* in both its native (Japanese) and invasive (British and French) ranges. We describe 18 new RNA viruses, representing 10 different virus families, and confirm their presence in RNA pools using RT-PCR. We place these viruses in the phylogenetic context of recent metatransciptomic studies in the host genus (Webster et al., 2015, Webster et al., 2016) and in invertebrates as a whole (Shi et al., 2016).

Methods

Sample collection

In total, we collected approximately 4450 individual *D. suzukii* across a three-year period between September 2013 and September 2016, including 230 larvae in 2016. Flies were collected near Montpellier, France (43.59 N, 3.78 E) in 2013 (collection by AX and SF); in Kent, UK (51.284 N, 0.465 E) during the late summer of 2014, 2015 and 2016 (NCM); and in three locations across Honshu, Japan, during May 2016 (NCM and MN): Tokyo University of Agriculture and Technology, Fuchu (35.683 N, 139.481 E); Naganuma Park, Tokyo (35.637 N, 139.375 E); Shimaminami Shima, Yamagata Prefecture (38.351 N, 140.276 E); Agriculture Total Center Kaju Research Institute, Fukushima (37.813 N, 140.443 E); and Fuefukigawa Fruit Park, Yamanashi (35.700 N, 138.666 E). We used a combination of commercial bait traps with cotton soaked in a proprietary liquid attractant (DROSO TRAP® and DROS'ATTRACT®, Biobest, Belgium, NV), and a standard sweep net to catch adult flies. Traps, hung at field margin and woodland sites, were collected at intervals of two to three days. All individuals were sorted into vials by trap and species within three hours of collection. We aimed to morphologically identify all species of *Drosophila* caught (Bächli et al., 2004), however, we also subsequently examined RNA pools for potential contamination due to misidentification. Other species of *Drosophila* were caught in these traps and we collected them together with *D. suzukii*, but they were not analysed further. Wild-collected flies were maintained on solid agar/sugar medium, before being macerated in sterile Ringer's solution (to allow for future experimental virus culture and isolation). Larvae were dissected from infested fruit collected in 2016 from UK and Japan with sterile forceps.

We immediately extracted RNA from a subsample of the fly (or larva) homogenate using TRIzol® (Invitrogen), before storage at -80°C. We treated pooled RNA samples for possible DNA

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

contamination using DNAse (Turbo DNA-free, Ambion) prior to library preparation. For flies collected in the UK and Japan, library preparation and sequencing was performed by Edinburgh Genomics (Edinburgh, UK) using the Illumina Hi-Seq platform with 120 or150nt paired end reads. To increase representation of viral and host protein coding RNAs, all libraries underwent depletion of rRNA using Ribo-Zero rRNA Removal Kit (Illumina). Flies collected in France during 2013 were sequenced separately at Beijing Genomics Institute (BGI tech solutions, Hong Kong) using paired-end 90nt reads. These samples underwent Duplex-Specific Thermostable Nuclease (DSN) normalisation and poly-A selection. This process, although enriching for viruses by rRNA depletion, biases virus discovery towards poly-adenylated genomic products only produced by certain viral taxa (e.g. Picornavirales). All raw reads have been submitted to the NCBI sequence read archive under project accession PRJNA402011 (Japan SRR6019484; France SRR6019487; Kent: SRR6019485, SRR6019486, and SRR6019488).

Virus identification and Phylogenetic Analysis

To remove those reads derived from *Drosophila*, we mapped raw reads against the *D. suzukii* genome and transcriptome using Bowtie2 (Langmead and Salzberg, 2012) with the '--very-fast' command-line option. We used Trimmomatic (Bolger et al., 2014) to quality trim and remove adapter sequences from the remaining unmapped raw reads (as pairs) using default parameters, before *de novo* assembly using Trinity version 2.2.0 (Grabherr et al., 2011), retaining a minimum contig length of 500nt. All raw unannotated contigs are provided in supporting file S.1_Data. We concatenated all translations of all open reading frames (ORFs) in each resulting contig, and retained only those with an open reading frame of 150 codons or greater. These concatenated protein sequences were used to search against a custom database using Diamond (Buchfink et al., 2015) with a an e-value threshold of 0.01, retaining a single top hit. This database comprised all of the viral proteins from the Genbank non-redundant protein database ('nr'; Clark et al., 2016), and all of the prokaryote, protist, fungal, nematode, hymenopteran, and dipteran sequences from NCBI refseq protein, Contigs for which the top hit was a virus were imported into Geneious®8.0.2 sequence analysis software (Kearse et al., 2012) for manual analysis. We grouped putative virus fragments taxonomically according to their initial best blast hit, assembled (Geneious) and manually curated them with reference to closest relatives in Genbank, to give the longest viral sequences consistent with the predicted protein content and structure of that virus taxon.

To infer phylogenetic relationships, we used RNA-dependent RNA polymerase (RdRp) coding sequences unless otherwise stated. The RdRp is generally the most conserved protein across RNA viruses, making it suitable for phylogenetic analysis of this diverse set of virus taxa (Koonin et al., 1993, Shi et al., 2016). RdRp gene sequences were translated and aligned with homologous sequences from their close relatives, as identified by BLAST (Altschul et al., 1990). To align multiple protein sequences we used ClustalW (Thompson et al., 2002) with BLOSSOM cost matrix (Henikoff and Henikoff, 1992). We manually identified regions of poor alignment at the 5' and 3' ends of the alignment and removed them before further analysis. All alignments are provided in supplementary material S2_Data. We then inferred maximum-likelihood phylogenetic trees using PhyML 2.2.3 (Guindon and Gascuel, 2003) with the LG substitution model (Le and Gascuel, 2008). We calculated branch support using the Shimodaira-Hasegawa-like nonparametric version of an

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

approximate likelihood ratio test implemented in PhyML (aLRT; Anisimova et al.,2011). Trees in Figures are clusters from within of larger trees (S3_Data), realigned and reconstructed using the same methods.

Detection by RT-PCR

To confirm the presence of the newly discovered viruses in original RNA pools we used Reverse Transcription PCR (RT- PCR) to screen for short amplicons of each virus' longest ORF, where possible spanning part of the RdRp gene. We designed primers using the Primer3 (Rozen and Skaletsky, 1999) plugin for Geneious (Kearse et al., 2012). RNA virus sequences identified by metagenomic methods may derive from viral elements endogenised into genomic DNA, if they are expressed (Katzourakis and Gifford, 2010). To test for endogenised viral elements (EVEs) we conducted PCRs (without a reverse transcription step) on RNA samples that contained genomic DNA from the original phenol-chloroform extraction. As RNA viruses do not produce a DNA intermediate, any viruses detected in this way are likely to be EVEs.

Virus Genome Annotation

For viruses with complete, or near complete genomes, we were able to infer genome structure and identify protein functional domains by first identifying ORFs and then comparing these to the Conserved Domain Database with an expected value threshold of 5×10^{-3}, and searching the NCBI 'nr' protein database using BLASTp. Only ORFs of 100 amino acids or longer were annotated, unless notable similarity to closely related viruses was evident. ORFs of less than 200 amino acids that were nested completely with larger ORFs were disregarded, unless they displayed high similarity to known proteins.

Distribution of RNA sequence Reads Across Samples

To estimate the number of virus reads in each pooled sample, and to detect any cross-species contamination in fly collections, we mapped trimmed forward reads to all new and previously published *Drosophila* virus genomes (including multiple divergent isolates where they were available), a selection of *Drosophila* ribosomal sequences, and a short region of cytochrome oxidase 1 (COI) that has discriminatory power between *Drosophila* species. Sequences were mapped with Bowtie2 (Langmead and Salzberg, 2012) using the '--very-sensitive' option. We report these after normalisation by the number of non-ribosomal reads and the length of each target sequence. We also apply an arbitrary lowest level detection threshold for each putative species of 0.5 total reads per Kb per million non-rRNA reads to reduce spurious signals caused by low level species contamination, library barcode switching, and cross-mapping to close relatives.

Results

In total, we generated approximately 280 million read pairs, ranging from 33 million pairs (Kent 2016) to 105 million pairs (France 2013) per library. Our assemblies comprised between 18,431 (Japan 2016) and 56,384 (Kent 2015) putative transcript contigs. Among these, we identified 18

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

new RNA viruses associated with *D. suzukii* (Table 1.). These viruses represent a variety of RNA virus taxa with positive sense single stranded (+ssRNA), negative sense single stranded (-ssRNA), and double stranded RNA (dsRNA) genomes, and include representatives of the Picornavirales, Mononegavirales, Bunyavirales, Chuviruses, Nodaviridae, Tombusviridae, Reoviridae and Nidovirales. We did not identify any DNA viruses. We do not report as new any viruses detected in *D. suzukii* that are identical, or near identical (>95% amino acid similarity in the polymerase), to previously published viruses. Those previously described viruses that were detected in *D. suzukii* are detailed in supplementary material (S1_Table.) and relative read counts in each pool are shown in Figure 6.

We have provisionally named these viruses according to the location from which the hosts were sampled. We have chosen not to include taxonomic or host information in the provisional name of the virus, as these are subject to change as phylogenetic relationships are revised and alternative or additional hosts discovered. The one exception to this rule is *D. suzukii* Nora Virus. This virus is sufficiently closely related to the *D. melanogaster* Nora virus and *D. immigrans* Nora virus that a name outside of this local scheme may cause confusion for future studies. During Phylogenetic analysis, a number of virus-like sequences were identified by BLAST in the public Transcriptome Shotgun Database (TSA). These have been included in analyses to improve accuracy of phylogenetic inference, but are not further discussed.

From our sample pools we detected reads mapping to a number of known *Drosophila* viruses. Because of the presence of closely related viruses in *D. suzukii* and other *Drosophila* species there is a possibility that reads of true *D. suzukii* viruses may cross map to their close relatives in other fly species. In addition, COI read mapping suggests a low level of species contamination (notably *D. melanogaster* and *D. immigrans*) in some pools.

Viruses with single-stranded positive sense RNA genomes.

Ten of the viruses described here encode their genomes in +ssRNA. Of these, Teise virus was found at the highest titre. Teise virus is a sobemo-like virus closely related to Prestney Burn Virus of *D. subobscura* (Webster et al., 2016) and Motts Mill virus of *D. melanogaster* (Webster et al., 2015), with 90.9% and 88.6% RdRp amino acid similarity respectively (Figure 1). The single-stranded positive sense genome of these viruses comprises two unjoined fragments, which may represent subgenomic products (Webster et al., 2015, Shi et al., 2016, Tokarz et al., 2014). However, *de novo* assembly of putative fragments of Teise virus produced contigs bridging the two fragments together, and PCR primers were designed that resultued in an amplicon bridging these fragment (see Fig. S1). We have therefore presented the genome of this virus as one contiguous fragment of 4.5 kb, while noting conserved domains (Fig. 2). Teise virus is the most geographically widespread virus of *D. suzukii*, with reads appearing in high numbers in both native and naturalised ranges (Fig. 6).

Medway virus (Fig. 1, B) shares close relationship to Braid Burn virus, previously described from *Drosophila* subsilvestris in the UK (Webster et al., 2016), . These viruses belong to a clade of insect viruses distantly related to the Sobemo and Polero viruses of plants. Medway virus appears at low copy-number in our samples with a small number of reads being detected in UK samples

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from 2014 and 2015. As for other viruses in this section of the Luteo-Sobemo groupthe Medway virus genome probably consists of two genomic RNA segments. However, the second RNA segment was not detected here and we describe the virus only from an RNA fragment that contains two ORFs, including the RdRp (Fig. 2). Tama virus, a third virus in the Luteo-Sobemo clade, was only detectable by PCR in Japanese samples.

In our *D. suzukii* collections we detected three separate Nora viruses, *D. melanogaster* Nora Virus (Habayeb et al., 2006), *D. immigrans* Nora Virus (van Mierlo et al., 2014) and the new Nora virus, most closely related to that of *D. immigrans*, but sufficiently divergent from both (37.1% and 30.4% amino-acid divergence at the RdRp locus, respectively) to merit description (Fig. 1, J). This clade of viruses also evidently infects other families of 'fruit fly', as they are detectably in the transcriptomes of two species of tephritids (*Bactrocera latifrons* and *Ceratitis capitata*), and can also be found in the transcriptomes of their parasitoid, *Fopius arisanus* (Fig. 1, J).

Beult virus was the most geographically widespread virus we identified: we detected Beult virus across sampling locations and years, with reads being especially abundant in samples from the UK in 2014 and Japan in 2016. Belonging to a clade of Virga-like viruses, it is very closely related to Bofa virus and Buckhurst virus of *D. melanogaster* and *D. obscura*, respectively (Webster et al., 2016). We identified two different haplotypes of this virus, which share a 98.9% nucleotide similarity: one from the UK, and a second divergent lineage from Japan. Saiwaicho virus also belongs to the Hepe-Virga clade of viruses and is closely related to the Negevirus genus. For the Hepe-Virga clade viruses we were able to identify domains for transferases, helicases, and polymerases (Fig. 2).

We detected a single Nidovirus in our samples from the UK and Japan. We have provisionally named this Fuefuki virus, and it has the longest contig recovered for any of our putative viruses, at over 16.5 kb. Within this near-complete genome we identify five ORFs but only one conserved domain: the RdRP (Fig. 2). Fuefuki virus is very closely related to Wuhan nido-like virus 1 (Shi et al., 2016) at 94.8% amino acid similarity in the polymerase. Along with Hubei *Tetragnatha maxillosa* virus 7 and Wuhan insect virus 19 (Shi et al., 2016) these four viruses form a distinct cluster near to the Coronaviridae, a family containing some notable vertebrate pathogens, including the SARS virus (Fig. 1, G).

Viruses with single-stranded negative sense RNA genomes

Five of the viruses we detected have -ssRNA genomes. Three of these belong to the Bunya-Arena clade of viruses: Notori virus, Ditton virus and Barming virus. Notori and Ditton viruses can be further classified as Phasmaviruses. These were detected in our samples as contigs of around 7kb in length that represent complete, or near-complete L- segments (Bishop and Shope, 1979) (Fig. 5). Barming virus, the third putative Bunya-Arena clade virus we identified, belongs to the Phlebo-like cluster of the clade. It too is known from a contig of just over 6kb, also representing the L-segment of the Bunyavirus genome, consisting of one ORF containing the viral RdRp (Fig. 4). The closest relative of Barming virus was a viral-like sequence identified in the TSA database from *Colletotrichum cereale*, a plant disease that has been found to cause crown rot anthracnose of turf grass (Crouch et al., 2006).

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The remaining -ssRNA viruses we identified belong to the Mono-Chu clade of -ssRNA viruses. From fly samples collected in the UK in 2014 we identified Kiln Barn virus, represented by a 3.7 kb contig. This clusters phylogenetically with a group of viruses close to the Chuviruses *sensu stricto,* and we find its closest relatives to be Hubei rhabdo-like virus 4 (Shi et al., 2016) and a viral sequence identified in the transcriptome of the Shiitake mushroom fungus *Lentinula edodes* (AGH07920.1). The other virus we identified from this clade, Mogami virus, is closely related to Shayang fly virus 1, a Chuvirus detected in Chinese diptera (Shi et al., 2016), and was represented by a 10.5kb contig in which from which we are able to identify both glycoprotein and polymerase ORFs.

Viruses with double-stranded RNA genomes.

We discovered three viruses predicted to possess double-stranded RNA genomes. These included two Totiviruses, Snodland virus and Larkfield virus, both reported from partial protein coding sequences. Both have closest relatives discovered in insect pool sequencing by Shi et al. (2016). Larkfield shares a cluster within the Totiviruses which includes a number of ant viruses: two discovered by Koyama et al. (2015) and Koyama et al. (2016) in genus *Camponotus*, and one found here as a virus like sequence in a published transcriptome of the black garden ant: *Lasius niger* (Fig. 5). Its closest relative, Hubei toti-like virus 14, is described as an endogenous viral element (Shi et al., 2016). Snodland virus clusters with a small group of other insect viruses, neighbouring a cluster of mycoviruses associated primarily with powdery mildews (Fig. 5).

The final dsRNA virus identified, Eccles virus, is our only representative of a virus family that has been previously advocated for the biological control of insect pests (Peng et al., 2000): the Reoviridae. Eccles virus is most closely related to Hubei diptera virus 21 (Shi et al., 2016) and a reovirus of the geometrid, *Operophtera brumata* (Graham et al., 2006). Homology predicts this virus has a multipartite genome consisting of 11 segments, although we were only able to assemble 6 of those segments from our samples.

Known Drosophila Viruses

We also detected 18 further viruses previously described from other species of *Drosophila*. Three known viruses were detected at very high levels (below), and are therefore highly likely to represent infections of *D. suzukii.* The first of these is Brandeis virus (MF953177) the genome of which is reported here for the first time. Although originally detected by Webster et al. (2015) in public *D. melanogaster* transcriptome datasets (PRJNA159179; Rodriguez et. al., 2012) and provisionally named, it has not previously been detected in wild flies. It is detected here at high levels (26.8% of all remapped virus reads) in *D. suzukii* samples from France in 2013. Brandeis virus belongs to the Hepe-Virga clade of +ssRNA viruses and is closely related to Muthill virus, a virus of *Drosophila*: *D. immigrans* (Webster et al., 2016). We were able to assemble a contig of 10.7 kb, which given homology to closely related virga-like viruses, is likely to represent a near-complete genome. Detection of Muthill virus itself is likely due to cross mapping of reads to this Brandeis virus genome. The other previously reported *Drosophila* viruses that we reidentified

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with confidence here are the iflaviruses Kinkell virus and La Jolla virus. Kinkell virus, first described by Webster et al. (2016) was detectable in *D. suzukii* from the UK in 2016, and La Jolla in all samples from all locations. La Jolla virus reads were detected at high titre in all our samples, comprising up to 30.7% of viral reads in British flies from 2014, and on average 15.0% of virus reads across all samples.

Four viruses of other *Drosophila* species also appear to be present in *D. suzukii* populations. For example, Corseley virus, a virus most associated with *D. subobscura* (Webster et al., 2016), which was detected at fairly high levels in British caught *D. suzukii* from 2016. It is uncommon in other *Drosophila* species (Webster et al., 2016) and is sufficiently divergent from any newly described *D. suzukii* viruses to minimise cross-mapping of reads. Galbut and Chaq viruses are both known infectious agents of *D. melanogaster*, but appear to be at high levels in 2015 *D. suzukii*. Cross-mapping to these viruses is unlikely due to their divergence from other *Drosophila* viruses, and host species contamination is unlikely to explain the high numbers of re-mapped reads observed. Dkikkawai virus (Webster et al., 2015) may represent true association for the same reasons. It was detected in Japanese flies only, although not at such high titre. Bloomfield virus, a reovirus of *D. melanogaster*, also likely represents true association with *D. suzukii* as we identified a divergent haplotype of one of the 10 genomic segments in *D. suzukii* that has not previously been seen in *D. melanogaster.* It is tempting to speculate that this reflects a history of host shifting and segment reassortment in this virus.

The remaining previously published viruses, were detected at much lower levels, and it is likely that some of them represent a low level of cross mapping from newly described but closely related viruses. For example, the low number of reads mapping to Buckhurst virus, a virus of *D. obscura* (Webster et al., 2016), may be mismapped Beult virus reads, the few reads mapping to Prestney Burn virus and Motts Mill (Webster et al., 2016, Webster et al., 2015) are most likely mismapped Teise virus reads. Reads mapping to *D. immigrans* Nora virus and *D. melanogaster* Nora virus may actually belong to the *D. suzukii* Nora virus. In addition, a small number may result from sample contamination by misidentified flies and/or library cross-contamination (such as barcode-switching, see: Sinha et al. 2017; Kircher et al., 2011; and Ballenghein et al., 2017). This includes viruses with no close relative associated with *D. suzukii*, such as Thika virus, Craigies Hill virus and Ashworth virus (*unpublished*), or viruses with biologically constrained host ranges, such as the Sigma viruses, along with *Drosophila* A virus (DAV), Drosophila C virus (DCV), and *D. melanogaster* Nora virus that were known to be present in *D. melanogaster* samples run alongside the 2016 *D. suzukii* samples.

Virus titre and composition varies among samples

To estimate the amount of virus in each of our samples we mapped all raw reads back to newly discovered and established putative *Drosophila* virus genomes (Fig. 6). The percentage of non-rRNA reads that mapped to any *Drosophila* virus varied from 0.09% in the poly-A selected French sample up to 5.14% in UK sample from 2016, with an average of 4.27% of reads being viral in Japanese and British pools. The virus composition varied markedly among samples from different times and locations. Despite applying a detection threshold for very low viral read numbers, it is not possible to formally analyse patterns of virus sharing among years or sampling locations from these data, for three reasons: First, the possibility of cross-mapping between some highly-

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conserved regions of the most closely related viruses means that we cannot accurately infer presence and absence of those viruses. Second, barcode switching (Sinha et al., 2017, Kircher et al., 2011, Ballenghien et al., 2017) and other sources of cross-contamination between libraries sequenced together on the Illumina platform may allow miss-assignment of reads between the Japanese and British samples from 2016, and also from other drosophilid libraries analysed at the same time. Finally, cytochrome oxidase read mapping suggests a small proportion of contaminating reads deriving from *D. melanogaster* and *D. immigrans* were present in some of our datasets. For example, in the Japanese sample of 2016 1.3% of COI reads mapped to *D. immigrans* (potentially misidentified larvae) and in the UK sample of 2015 0.74% of reads mapped to *D. melanogaster*. The *D. melanogaster* reads may represent misidentification or cross-mapping, as the species are quite closely related, but it is more likely that they are the result of contamination across libraries through barcode switching as *D. suzukii* samples were sequenced in parallel with unrelated drosophilid libraries.

Discussion

Here we make a first survey of the viruses associated with the invasive *Drosophila* pest *D. suzukii* in its native and invasive ranges. Alongside 18 new viruses, not previously described from any organism, we confidently identified a further seven viruses associated with this novel invasive host that had previously been described from other *Drosophila* species. Some novel viruses were detected solely from the native range of *D. suzukii* and others from the invasive range, but rarely from both habitats.

These viruses were identified from metagenomic sequencing of samples of wild *D. suzukii*, and although their presence as RNA but not DNA implies that they are not expressed endogenised viral elements (i.e. EVEs), it remains possible that some are not truly infections in this fly, but may be contaminants of the surface of the fly or infect a commensal, pathogenic, or food organism within the fly's gut lumen. However, we believe that this is unlikely to be the case for most sequences, as previous studies that additionally used the presence of virus-derived 21nt short interfering RNAs to demonstrate active replication (Webster et al., 2015) found that the majority of viruses identified in similar metatransciptomic sequencing of *D. melanogaster* constituted active infections. There is also a possibility of some cross-species contamination, barcode-switching and cross-mapping could result in spurious host allocation but this is not compatible with the read number or distribution of the majority of viruses (above).

In addition, more recent large-scale invertebrate virus discovery projects (Shi et al., 2016) give us a greatly increased confidence in the phylogenetic relationships of newly identified virus sequences. In particular, despite some virus taxa having a diverse host ranges, it seems reasonable to infer that *D. suzukii* is the true host for viruses with very close relatives confirmed to infect another insect. For example, Mogami virus (Chuvirus) is distantly related to any known *Drosophila* virus, but is closely related to Shayang Fly virus 1 (Shi et al., 2016) and clusters within a group of viruses that are only described from insect samples (see Fig. 4, D). Nevertheless, this pattern is not true for all viruses described here. Specifically, two of the 18 novel viruses in this study (Ditton virus and Barming virus), are more closely related to Mycoviruses than they are to

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

any entomopathogenic viruses and one(Luckshill virus) is most closely related to a sequence found in a parasitic nematode of ruminants. And, while this pattern does not exclude the possibility of these being true viruses of *D. suzukii*—as many viral families contain a broad range of hosts including those of different phyla and patterns of host switching are still little understood—these are among the best candidates to be infections of *Drosophila* parasites or gut fauna, rather than *D. suzukii* itself.

The potential of these viruses to be used as biological control agents is currently unclear. Commercially successful viral biocontrol agents have in the past only come from the dsDNA virus family Baculoviridae, which was not represented in our collections, and most lineages represented here have not been investigated for their ability to be cultured and applied as control agents. Indeed, few viruses in the families here have been successfully isolated for experimentation, and many are known only from metagenomic sequencing. The only virus family we found in associated with *D. suzukii* that has any history as a control agent (Zeddam et al., 2003, Peng et al., 1998, Peng et al., 2000)is the reovirus 'Eccles virus'. Eccles virus was relatively rare in our samples, but this may speak to the potential pathogenicity of the virus, as flies harbouring a particularly pathogenic virus, especially one that has a short latency period, may be less likely to visit baited traps (Gupta et al., 2017). Further investigation of this virus, including isolation and pathogenicity assays, are needed before any further conclusions can be drawn about its utility as a control agent. Viruses potentially lethal to *D. suzukii* may also await discovery in other species of *Drosophila.* Indeed, pathogens have the potential to display increased virulence following a host shift event (Longdon et al., 2015) and the susceptibility of *D. suzukii* to viruses of *D. melanogaster* has been shown experimentally (Cattel et al., 2016, Lee and Vilcinskas, 2017). Here we show the potential association of viruses from *D. melanogaster*, *D. immigrans* and *D. subobscura* with *D. suzukii* in the wild. Further investigation of the virological environment experienced by many different *Drosophila* in nature may, therefore, be of both academic and applied interest.

Given our focus on an invasive species, the potential for a shift in the virological environment associated with invasion is of particular interest. Theory predicts that organisms may experience a 'release' from natural enemies, including pathogens, in their invasive range (Keane and Crawley, 2002) but this idea remains contentious, as supporting evidence is limited (Colautti et al., 2004). It has also been hypothesised that invasives, rather than experience a drop in overall number of enemies, undergo a shift in the type of enemy encountered, from co-evolved specialists in the native range to more generalist enemies, quickly able to adapt to a new host, in the naturalized range (Joshi and Vrieling, 2005). In this study, we do detect a an apparently marked difference in the virus communities of flies from different areas within its expanding geographical range. Although the potential for cross-mapping, and a low level of species contamination, mean that these findings should be treated with some caution, five of the new viruses described (Saiwaicho virus, Tama virus, Mogami virus, Nagamura virus and Notori virus) were only detected at high levels in Japanese (native) flies. These five viruses are not particularly closely related to any previously described *Drosophila* viruses (Fig. 1 and Fig. 3) and may represent a more specialized relationship with *D. suzukii*. In contrast, the three most ubiquitous viruses across all samples, La Jolla virus, Teise virus and Beult virus are a either a known generalist (La Jolla) or very closely related to a virus in another related hosts (Teise and Beult; Fig. 1). If

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

confirmed, this pattern could reflect a shift in natural enemy type from native to invasive range of *D. suzukii*.

Acknowledgements

We thank Michelle Fountain, Bethan Shaw, Adrian Harris, Maddie Cannon, and the entomological team at East Malling research for support in collecting British flies. We thank members of the Nakai lab at the Tokyo University of Agriculture and Technology for making us welcome, and the staff at the TUAT Field Museum Tama hills, Yamagata Prefectural Agricultural Research Centre, Agriculture Center Kaju Institute, Fukushima and Yamanashi Prefecture Outside Organization Fruit Tree Test Site for support during field collections. We thank Edinburgh Genomics for sequencing, and members of the Obbard and Vale labs for discussion. This work was funded by a Wellcome Trust Research Career Development Fellowship (WT085064) to DJO and an Agricultural and Horticultural Development Board grant to DJO and JVC to support NCM's PhD. Work in DJO's laboratory was partly supported by a Wellcome Trust strategic award to the Centre for Immunity, Infection, and Evolution (WT095831).

References

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

AMIN UD DIN, M., MAZHAR, K., HAQUE, S. & AHMED, M. 2005. A preliminary report on Drosophila fauna of Islamabad (Capital, Pakistan). *Drosoph. Info. Serv.,* 88**,** 6-7.

ANISIMOVA, M., GIL, M., DUFAYARD, J. F., DESSIMOZ, C. & GASCUEL, O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol,* 60**,** 685-99.

ASPLEN, M. K., ANFORA, G., BIONDI, A., CHOI, D.-S., CHU, D., DAANE, K. M., GIBERT, P., GUTIERREZ, A. P., HOELMER, K. A. & HUTCHISON, W. D. 2015. Invasion biology of spotted wing Drosophila (Drosophila suzukii): a global perspective and future priorities. *Journal of Pest Science,* 88**,** 469-494.

ATALLAH, J., TEIXEIRA, L., SALAZAR, R., ZARAGOZA, G. & KOPP, A. 2014. *The making of a pest: the evolution of a fruit-penetrating ovipositor in Drosophila suzukii and related species*.

BÄCHLI, G., VILELA, C. R., ESCHER, S. A. & SAURA, A. 2004. *The Drosophilidae (Diptera) of Fennoscandia and Denmark*, Brill Academic Publishers.

BALLENGHIEN, M., FAIVRE, N. & GALTIER, N. 2017. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC biology,* 15**,** 25.

BERGOIN, M. & TIJSSEN, P. 1998. Biological and Molecular Properties of Densoviruses and Their Use in Protein Expression and Biological Control. *In:* MILLER, L. K. & BALL, L. A. (eds.) *The Insect Viruses.* Boston, MA: Springer US.

BISHOP, D. H. L., ENTWISTLE, P. F., CAMERON, I. R., ALLEN, C. J. & POSSEE, R. D. 1988. Field trials of genetically-engineered baculovirus insecticides.

BISHOP, D. H. L. & SHOPE, R. E. 1979. Bunyaviridae. *In:* FRAENKEL-CONRAT, H. & WAGNER, R. R. (eds.) *Comprehensive Virology: Newly Characterized Vertebrate Viruses.* Boston, MA: Springer US.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BOLDA, M. 2008. *New fruit fly pest in strawberries and caneberries* [Online]. University of California, Division of Agriculture and Natural Resources. Available: http://ucanr.edu/blogs/blogcore/postdetail.cfm?postnum=821 [Accessed 10/11/2014 2014].

BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120.

BRONKHORST, A. W. & VAN RIJ, R. P. 2014. The long and short of antiviral defense: small RNA-based immunity in insects. *Current Opinion in Virology,* 7**,** 19-28.

BUCHFINK, B., XIE, C. & HUSON, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Meth,* 12**,** 59-60.

CALABRIA, G., MÁCA, J., BÄCHLI, G., SERRA, L. & PASCUAL, M. 2012. First records of the potential pest species Drosophila suzukii (Diptera: Drosophilidae) in Europe. *Journal of Applied entomology,* 136**,** 139-147.

CATTEL, J., MARTINEZ, J., JIGGINS, F., MOUTON, L. & GIBERT, P. 2016. Wolbachia-mediated protection against viruses in the invasive pest Drosophila suzukii. *Insect molecular biology,* 25**,** 595-603.

CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & SAYERS, E. W. 2016. GenBank. *Nucleic Acids Research,* 44**,** D67-D72.

COLAUTTI, R. I., RICCIARDI, A., GRIGOROVICH, I. A. & MACISAAC, H. J. 2004. Is invasion success explained by the enemy release hypothesis? *Ecology Letters,* 7**,** 721-733.

CROUCH, J. A., CLARKE, B. B. & HILLMAN, B. I. 2006. Unraveling Evolutionary Relationships Among the Divergent Lineages of Colletotrichum Causing Anthracnose Disease in Turfgrass and Corn. *Phytopathology,* 96**,** 46-60.

ELGEE, E. 1975. Persistence of a virus of the white-marked tussock moth on Balsam Fir foliage. *Bi-monthly Research Notes,* 31**,** 33-34.

FRAIMOUT, A., DEBAT, V., FELLOUS, S., HUFBAUER, R. A., FOUCAUD, J., PUDLO, P., MARIN, J.-M., PRICE, D. K., CATTEL, J. & CHEN, X. 2017. Deciphering the Routes of invasion of Drosophila suzukii by Means of ABC Random Forest. *Molecular biology and evolution,* 34**,** 980-996.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOHEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. & REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech,* 29**,** 644-652.

GRAHAM, R. I., RAO, S., POSSEE, R. D., SAIT, S. M., MERTENS, P. P. C. & HAILS, R. S. 2006. Detection and characterisation of three novel species of reovirus (Reoviridae), isolated from geographically separate populations of the winter moth Operophtera brumata (Lepidoptera: Geometridae) on Orkney. *Journal of invertebrate pathology,* 91**,** 79-87.

GUINDON, S. & GASCUEL, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology,* 52**,** 696-704.

GUPTA, V., STEWART, C., RUND, S. S. C., MONTEITH, K. & VALE, P. F. 2017. Costs and benefits of sub-lethal Drosophila C virus infection. *Journal of Evolutionary Biology*.

HABAYEB, M. S., EKENGREN, S. K. & HULTMARK, D. 2006. Nora virus, a persistent virus in Drosophila, defines a new picorna-like virus family. *Journal of General Virology,* 87**,** 3045-3051.

HARRIS, A. & SHAW, B. 2014. First record of Drosophila suzukii (Matsumura) (Diptera, Drosophilidae) in Great Britain. *Dipterists Digest,* 21**,** 189-192.

HENIKOFF, S. & HENIKOFF, J. G. 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America,* 89**,** 10915-10919.

HUGER, A. M. 2005. The Oryctes virus: its detection, identification, and implementation in biological control of the coconut palm rhinoceros beetle, Oryctes rhinoceros (Coleoptera: Scarabaeidae). *Journal of invertebrate pathology,* 89**,** 78-84.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

HUNTER-FUJITA, F. R., ENTWISTLE, P., EVANS, H. & CROOK, N. 1998. *Insect viruses and pest management*, John Wiley & Sons Ltd.

HUSZAR, T. & IMLER, J. L. 2008. Chapter 6 Drosophila Viruses and the Study of Antiviral Host-Defense. 72**,** 227-265.

JOSHI, J. & VRIELING, K. 2005. The enemy release and EICA hypothesis revisited: incorporating the fundamental difference between specialist and generalist herbivores. *Ecology Letters,* 8**,** 704-714.

KANESHIRO, K. Drosophila (Sophophora) suzukii (Matsumura). Proc Hawaiian Entomol Soc, 1983. 179.

KANG, Y. & MOON, K. 1968. Drosophilid fauna of six regions near the demilitarized zone in Korea. *Korean Journal of Zoology,* 11**,** 65-68.

KANZAWA, T. 1935. Research into the fruit-fly Drosophila suzukii Matsumura (preliminary report). *Yamanashi Prefecture Agricultural Experiment Station Report*.

KANZAWA, T. 1939. Studies on Drosophila suzukii Mats. *Studies on Drosophila suzukii Mats.*

KARAGEORGI, M., BRÄCKER, L. B., LEBRETON, S., MINERVINO, C., CAVEY, M., SIJU, K. P., KADOW, I. C. G., GOMPEL, N. & PRUD'HOMME, B. 2017. Evolution of multiple sensory systems drives novel egg-laying behavior in the fruit pest Drosophila suzukii. *Current Biology,* 27**,** 847-853.

KATZOURAKIS, A. & GIFFORD, R. J. 2010. Endogenous viral elements in animal genomes. *PLoS genetics,* 6**,** e1001191.

KEANE, R. M. & CRAWLEY, M. J. 2002. Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution,* 17**,** 164-170.

KEARSE, M., MOIR, R., WILSON, A., STONES-HAVAS, S., CHEUNG, M., STURROCK, S., BUXTON, S., COOPER, A., MARKOWITZ, S., DURAN, C., THIERER, T., ASHTON, B., MEINTJES, P. & DRUMMOND, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28**,** 1647-1649.

KEMP, C. & IMLER, J. L. 2009. Antiviral immunity in drosophila. *Curr Opin Immunol,* 21**,** 3-9.

KIRCHER, M., SAWYER, S. & MEYER, M. 2011. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic acids research,* 40**,** e3-e3.

KOONIN, E. V., DOLJA, V. V. & MORRIS, T. J. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical reviews in biochemistry and molecular biology,* 28**,** 375-430.

KOYAMA, S., SAKAI, C., THOMAS, C. E., NUNOURA, T. & URAYAMA, S.-I. 2016. A new member of the family Totiviridae associated with arboreal ants (Camponotus nipponicus). *Archives of Virology,* 161**,** 2043-2045.

KOYAMA, S., URAYAMA, S.-I., OHMATSU, T., SASSA, Y., SAKAI, C., TAKATA, M., HAYASHI, S., NAGAI, M., FURUYA, T., MORIYAMA, H., SATOH, T., ONO, S.-I. & MIZUTANI, T. 2015. Identification, characterization and full-length sequence analysis of a novel dsRNA virus isolated from the arboreal ant Camponotus yamaokai. *Journal of General Virology,* 96**,** 1930-1937.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods,* 9**,** 357-359.

LE, S. Q. & GASCUEL, O. 2008. An improved general amino acid replacement matrix. *Molecular biology and evolution,* 25**,** 1307-1320.

LEE, K.-Z. & VILCINSKAS, A. 2017. Analysis of virus susceptibility in the invasive insect pest Drosophila suzukii. *Journal of Invertebrate Pathology,* 148**,** 138-141.

LIU, H. & STILING, P. 2006. Testing the enemy release hypothesis: a review and meta-analysis. *Biological Invasions,* 8**,** 1535-1545.

LONGDON, B., HADFIELD, J. D., DAY, J. P., SMITH, S. C. L., MCGONIGLE, J. E., COGNI, R., CAO, C. & JIGGINS, F. M. 2015. The Causes and Consequences of Changes in Virulence following Pathogen Host Shifts. *PLOS Pathogens,* 11**,** e1004728.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LUE, C.-H., L. MOTTERN, J., WALSH, G. C. & BUFFINGTON, M. L. 2017. New record for the invasive spotted wing drosophila, Drosophila suzukii (Matsumura, 1931)(Diptera: Drosophilidae) in Anillaco, western Argentina. *Proceedings of the Entomological Society of Washington,* 119**,** 146-150.

MATSUMURA, S. 1931. 6000 Illustrated Insects of Japan-Empire, Tokyo. *Sphyracephala quadriguttata (Walk.) figured***,** 371.

MITCHELL, C. E. & POWER, A. G. 2003. Release of invasive plants from fungal and viral pathogens. *Nature,* 421**,** 625-627.

MUSSABEKOVA, A., DAEFFLER, L. & IMLER, J.-L. 2017. Innate and intrinsic antiviral immunity in Drosophila. *Cellular and Molecular Life Sciences,* 74**,** 2039-2054.

OKADA, T. 1976. New distribution records of the drosophilids in the Oriental Region. *Acta Dipterologica,* 8**,** 1-8.

PARSHAD, R. & DUGGAL, K. 1965. Drosophilidae of Kashmir, India. *Drosophila Information Service,* 40**,** 44.

PENG, F. T. 1937. On some species of Drosophila from China. *Annotationes Zoologicae Japonenses,* 16**,** 20-27.

PENG, H.-Y., ZHOU, X.-M., SHENG, R.-J., CHEN, S.-W., LIU, J.-X., CHEN, X.-W., YANG, C.-H. & XIE, T.-E. 2000. Development of Dendrolimus punctatus wenshanensis Cytoplasm Polyhedrosis virus(DpwCPV) Insecticide. *Virologica Sinica,* 15**,** 148-155.

PENG, H., CHEN, X., JIANG, Y., SHEN, R., ZHOU, X. & HU, Z. 1998. Controlling Dendrolimus punctatus with Trichogramma dendrolimi carrying cytoplasmic polyhedrosis virus. *Chinese Journal of Biological Control,* 14**,** 111-114.

R CORE TEAM 2017. R: A Language and Environment for Statistical Computing.

ROZEN, S. & SKALETSKY, H. 1999. Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics methods and protocols***,** 365-386.

SABIN, L. R., HANNA, S. L. & CHERRY, S. 2010. Innate antiviral immunity in Drosophila. *Current opinion in immunology,* 22**,** 4-9.

SHI, M., LIN, X.-D., TIAN, J.-H., CHEN, L.-J., CHEN, X., LI, C.-X., QIN, X.-C., LI, J., CAO, J.-P., EDEN, J.-S., BUCHMANN, J., WANG, W., XU, J., HOLMES, E. C. & ZHANG, Y.-Z. 2016. Redefining the invertebrate RNA virosphere. *Nature,* advance online publication.

SHIEH, T. R. & BOHMFALK, G. T. 1980. Production and efficacy of baculviruses. *Biotechnology and Bioengineering,* 22**,** 1357-1375.

SIDORENKO, V. 1992. New and unrecorded species of Drosophilidae from Soviet Far East (Diptera, Brachycera). *Spixiana,* 15**,** 93-95.

SINHA, R., STANLEY, G., GULATI, G. S., EZRAN, C., TRAVAGLINI, K. J., WEI, E., CHAN, C. K. F., NABHAN, A. N., SU, T. & MORGANTI, R. M. 2017. Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *bioRxiv***,** 125724.

STERN, V., SMITH, R., VAN DEN BOSCH, R. & HAGEN, K. 1959. The integration of chemical and biological control of the spotted alfalfa aphid: the integrated control concept. *California Agriculture,* 29**,** 81-101.

SWOBODA-BHATTARAI, K. & BURRACK, H. Drosophila suzukii infestation in ripe and ripening caneberries. XI International Rubus and Ribes Symposium 1133, 2015. 419-430.

THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2002. Multiple Sequence Alignment Using ClustalW and ClustalX. *Current Protocols in Bioinformatics.* John Wiley & Sons, Inc.

TODA, M. 1991. Drosophilidae (Diptera) in Myanmar (Burma) VII. The Drosophila melanogaster species-group, excepting the D. montium species-subgroup. *Oriental Insects,* 25**,** 69-94.

TOKARZ, R., WILLIAMS, S. H., SAMEROFF, S., LEON, M. S., JAIN, K. & LIPKIN, W. I. 2014. Virome analysis of Amblyomma americanum, Dermacentor variabilis, and Ixodes scapularis ticks reveals novel highly divergent vertebrate and invertebrate viruses. *Journal of virology,* 88**,** 11480-11492.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

TORCHIN, M. E., LAFFERTY, K. D., DOBSON, A. P., MCKENZIE, V. J. & KURIS, A. M. 2003. Introduced species and their missing parasites. *Nature,* 421**,** 628-630.

VAN MIERLO, J. T., OVERHEUL, G. J., OBADIA, B., VAN CLEEF, K. W., WEBSTER, C. L., SALEH, M.-C., OBBARD, D. J. & VAN RIJ, R. P. 2014. Novel Drosophila viruses encode host-specific suppressors of RNAi. *PLoS pathogens,* 10**,** e1004256.

WALSH, D. B., BOLDA, M. P., GOODHUE, R. E., DREVES, A. J., LEE, J., BRUCK, D. J., WALTON, V. M., O'NEAL, S. D. & ZALOM, F. G. 2011. *Drosophila suzukii* (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding its Geographic Range and Damage Potential. *Journal of Integrated Pest Management,* 2**,** 1-7.

WARNES, G. R., BOLKER, B., BONEBAKKER, L., GENTLEMAN, R., HUBER, W. & LIAW, A. 2016. gplots: Various R Programming Tools for Plotting Data. 2016; R package version 3.0. 1. *Reference Source*.

WEBSTER, C. L., LONGDON, B., LEWIS, S. H. & OBBARD, D. 2016. Twenty five new viruses associated with the Drosophilidae (Diptera). *bioRxiv*.

WEBSTER, C. L., WALDRON, F. M., ROBERTSON, S., CROWSON, D., FERRARI, G., QUINTANA, J. F., BROUQUI, J.-M., BAYNE, E. H., LONGDON, B. & BUCK, A. H. 2015. The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. *PLoS Biol,* 13**,** e1002210.

XU, J. & CHERRY, S. 2014. Viruses and antiviral immunity in Drosophila. *Dev Comp Immunol,* 42**,** 67-84.

ZAMBON, R. A., VAKHARIA, V. N. & WU, L. P. 2006. RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster. *Cellular microbiology,* 8**,** 880-889.

ZEDDAM, J.-L., CRUZADO, J. A., RODRIGUEZ, J. L., RAVALLEC, M. & SUBILETE, E. C. 2003. A cypovirus from the South American oil-palm pest Norape argyrrhorea and its potential as a microbial control agent. *BioControl,* 48**,** 101-112.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Novel viruses detected in *D. suzukii*. [a]PCR reactions performed on cDNA. [b]PCR reactions performed on extractions containing nuclear DNA.

Provisional Name	Accession	Host	Taxon	Genome	Longest contig (kb)	Sample(s)	Detected by RT-PCR[a]	Detected by PCR[b]
Beult virus	MF893261, MF893262	Dsuz	Negevirus	+ssRNA	12	France2013, UK2014, UK2015, UK2016, Japan	+	-
Saiwaicho virus	MF893256	Dsuz	Negevirus	+ssRNA	10	Japan2016	+	-
Luckshil l virus	MF893250	Dsuz	Virgavirus	+ssRNA	3.5	UK2016	+	-
Tiese virus	MF893259	Dsuz	Luteoviridae	+ssRNA	4.5	France2013, UK2014, UK2015, UK2016, Japan2016	+	-
Tama virus	MF893258	Dsuz	Sobemovirus	+ssRNA	3.5	Japan2016	+	-
Medway virus	MF893251	Dsuz	Sobemovirus	+ssRNA	2.7	UK2014	+	-
Dsuz Nora virus	MF893254	Dsuz	Picornaviridae	+ssRNA	12	Japan2016	+	-
Naganuma virus	MF893253	Dsuz	Nodaviridae	+ssRNA	1.6	Japan2016	+	-
Fuefuki virus	MF893247	Dsuz	Nidoviridae	+ssRNA	16	Japan2016	+	-
Cyril virus	MF893263	Dsuz	Virgavirus	+ssRNA	1.5	UK2016	+	-
Eccles Virus	MF893265-MF893270	Dsuz	Reoviridae	dsRNA	4.2	UK2014	+	-
Larkfield virus	MF893249	Dsuz	Totiviridae	dsRNA	6	UK2015	+	-
Snodland virus	MF893257	Dsuz	Totiviridae	dsRNA	1.6	UK2015	+	-
Mogami virus	MF893252	Dsuz	Chuvirus	-ssRNA	10.5	Japan2016	+	-
Ditton virus	MF893264	Dsuz	Phasmaviridae	-ssRNA	7.3	UK2015	+	-
Barming virus	MF893260	Dsuz	Phleboviridae	-ssRNA	6.5	UK2016	+	-
Notori virus	MF893255	Dsuz	Phasmaviridae	-ssRNA	7	Japan2016	+	-
Kiln Barn virus	MF893248	Dsuz	Chuvirus	-ssRNA	3.7	UK2016	+	-

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

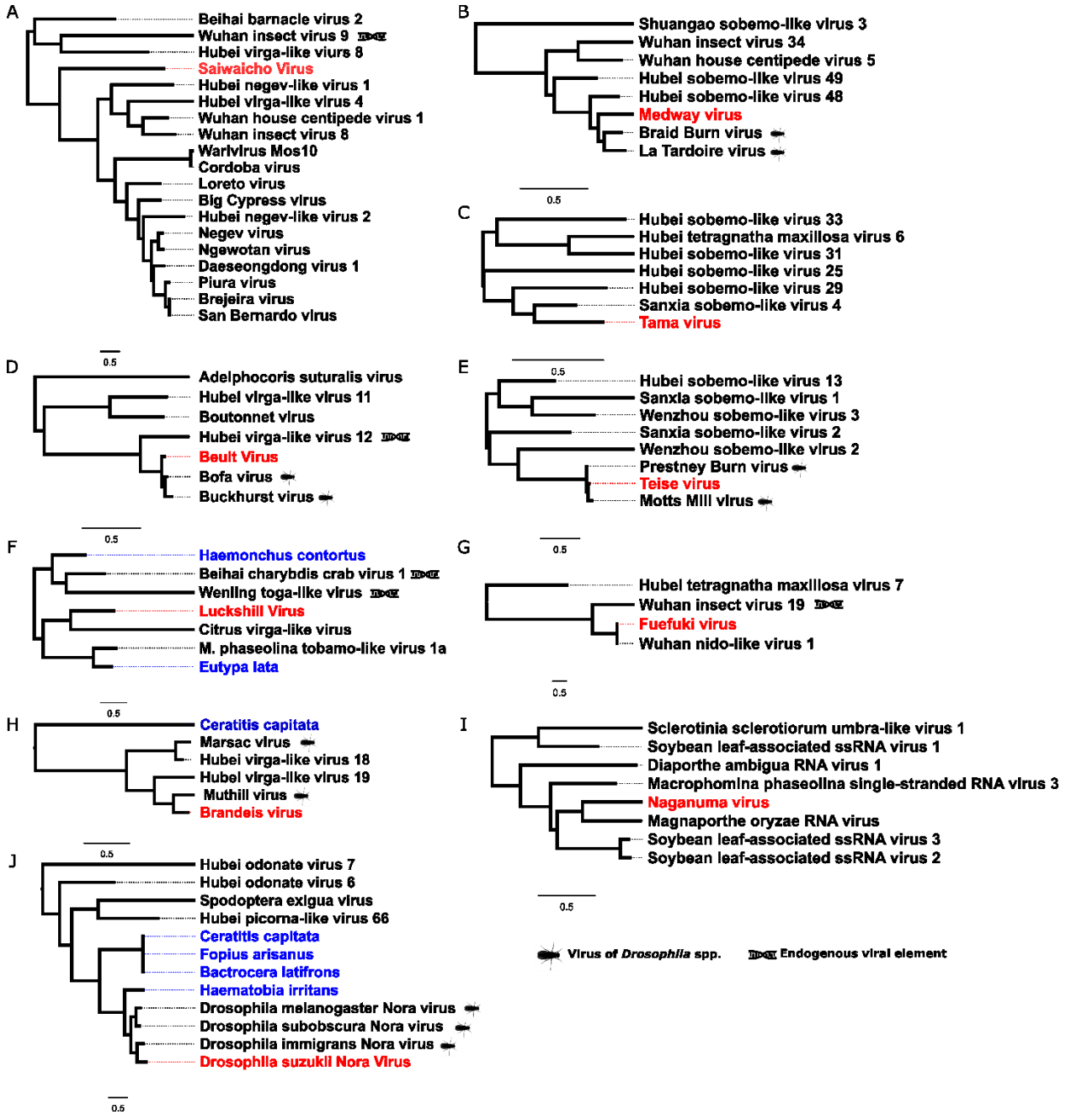


Figure 1. Positive sense single stranded RNA viruses. Midpoint-rooted, maximum-likelihood trees were inferred from viral polymerase sequences. Putative viruses newly described in association with *D. suzukii* (red) are highlighted alongside virus-like sequences identified in public transcriptome datasets (blue). Viruses previously described as endogenous viral elements are also marked. Tree **A:** Negeviruses and nearby clusters from the Hepe-Virga clade; **B,C,E:** Sobemo-like viruses belonging to clusters within the Luteo-Sobemo clade; **D:** Virga-like viruses belonging to the Hepe-Virga like clade ; **F:** A small cluster of toga-like viruses neighbouring the Alphaviruses, Togaviridae; **G:** a cluster in the Nidoviruses close to the Coronaviridae; **H:** Virga-like viruses belonging to the Hepe-Virga like clade; **I:** Cluster of Nodaviruses within the Tombus-Noda clade; **J:**Noraviruses and related cluster of the Picora-Calici clade.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

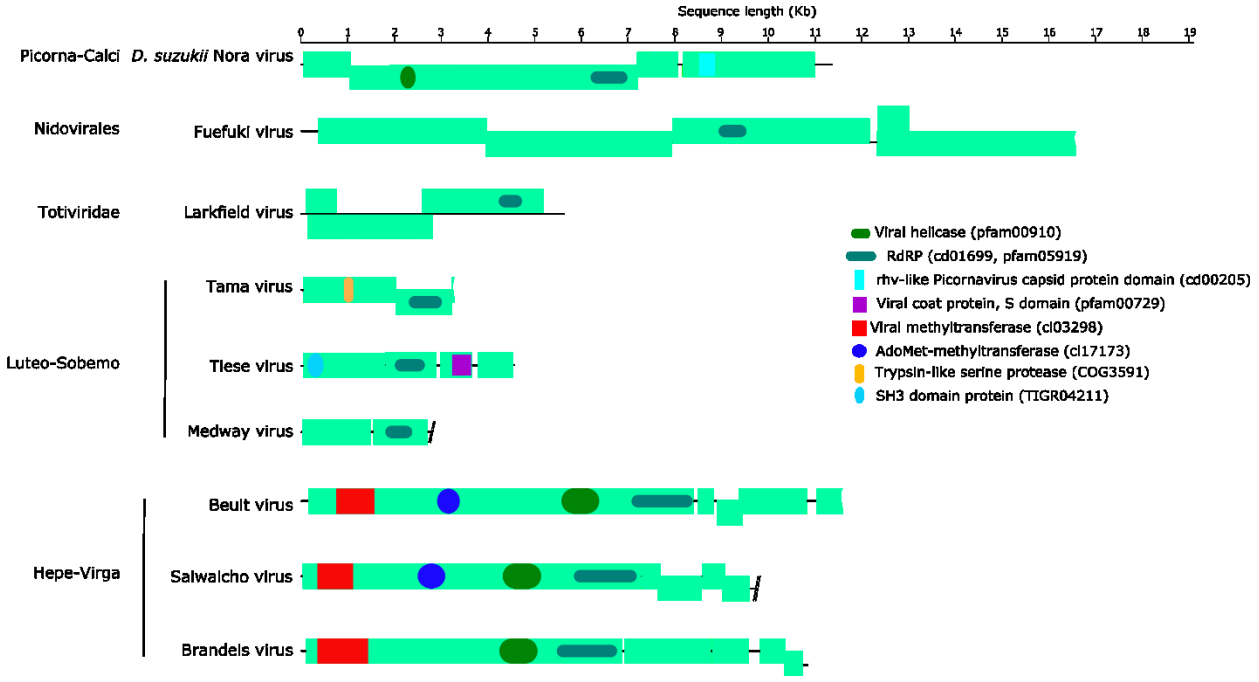


Figure 2. The structure of selected +ssRNA and dsRNA virus genomes for which we recover complete or near complete genome sequences. Relative positions of protein coding regions are calculated with reference to the NCBI conserved domain database.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

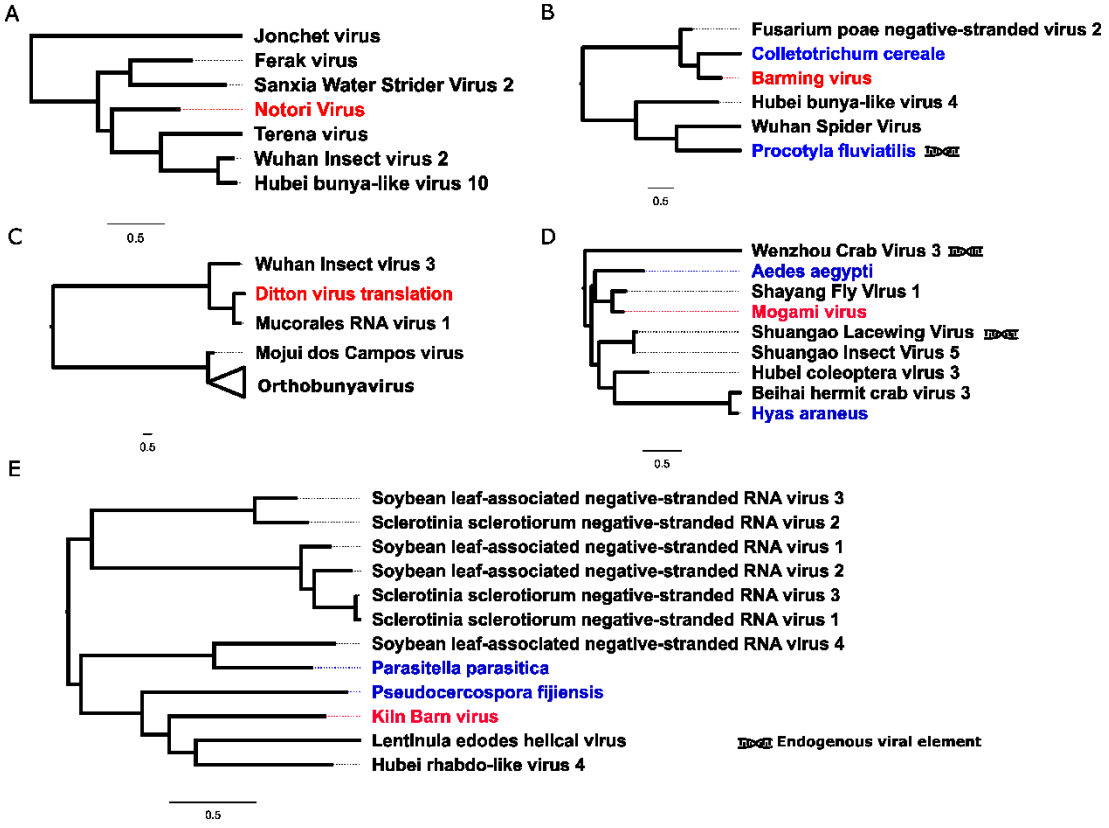


Figure 3. Negative sense single-stranded RNA viruses. Midpoint-rooted, maximum-likelihood trees were inferred from viral polymerase sequences. Viruses newly described in association with *D. suzukii* (red) are highlighted alongside viral-like sequences identified in public transcriptome datasets (blue). Viruses previously described by the original authors as endogenous viral elements are also marked. Tree **A:** Viruses close to Phasmaviruses in the Bunya-Arena group; **B:** Viruses belonging to the Phlebo-like cluster of the Bunya-Arena group; **C:** Orthobunyaviruses (collapsed) and small sister clade consisting of three viruses, including the newly described Ditton virus; **D:** Cluster of the Chuviruses; **E:** Cluster of viruses close to Chuviruses in the Mono-Chu clade.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

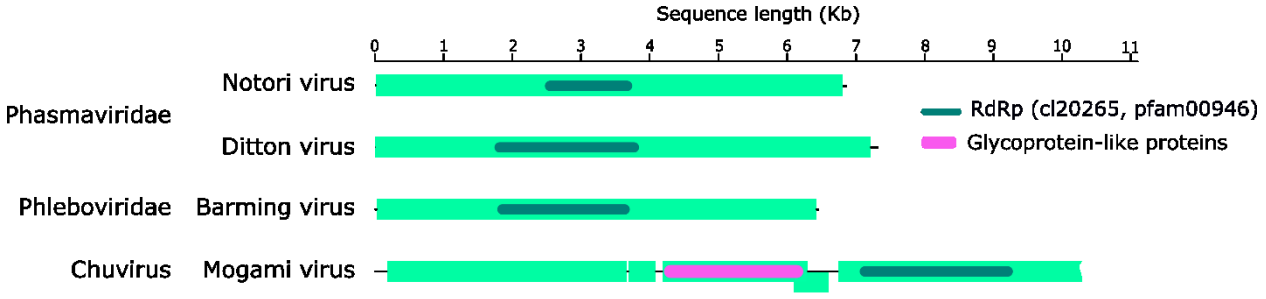


Figure 4. The structure of selected -ssRNA virus genomes for which we recover complete or near complete genome sequences. Relative positions of glycoprotein like proteins and RNA dependent RNA polymerase regions (RdRp) are calculated with reference to the NCBI conserved domain database.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

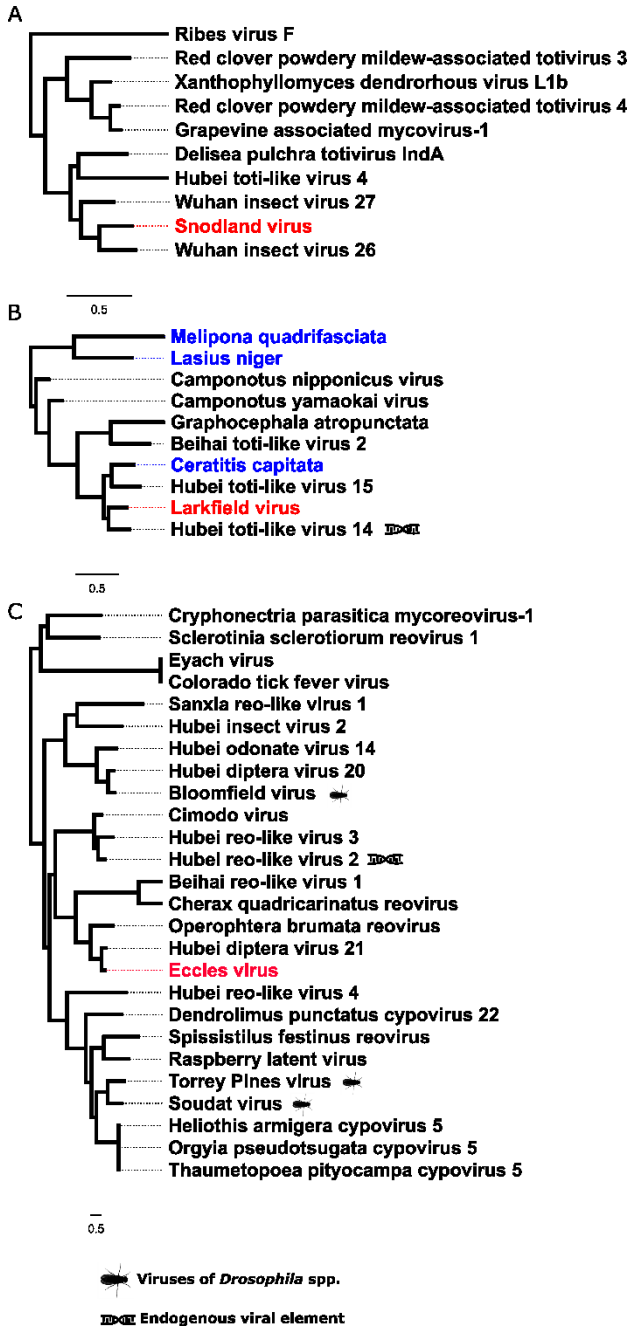


Figure 5. Double stranded RNA viruses. These midpoint-rooted, maximum-likelihood trees were inferred from viral polymerase sequences. Putative viruses newly described in association with *D. suzukii* (red) are highlighted alongside viral-like sequences identified in public transcriptome database (blue). Viruses previously described from a *Drosophila* spp. and viruses described by the original authors as endogenous viral elements are also marked. Tree **A:** Totiviruses, Totiviridae; **B:** Viruses belonging to a c;ade of the totiviridae, Toti-Chryso clade; **C:** Reoviruses, including Coltiviruses (Eyach virus and Colarado tick fever virus) and viruses close to Fijiviruses.

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

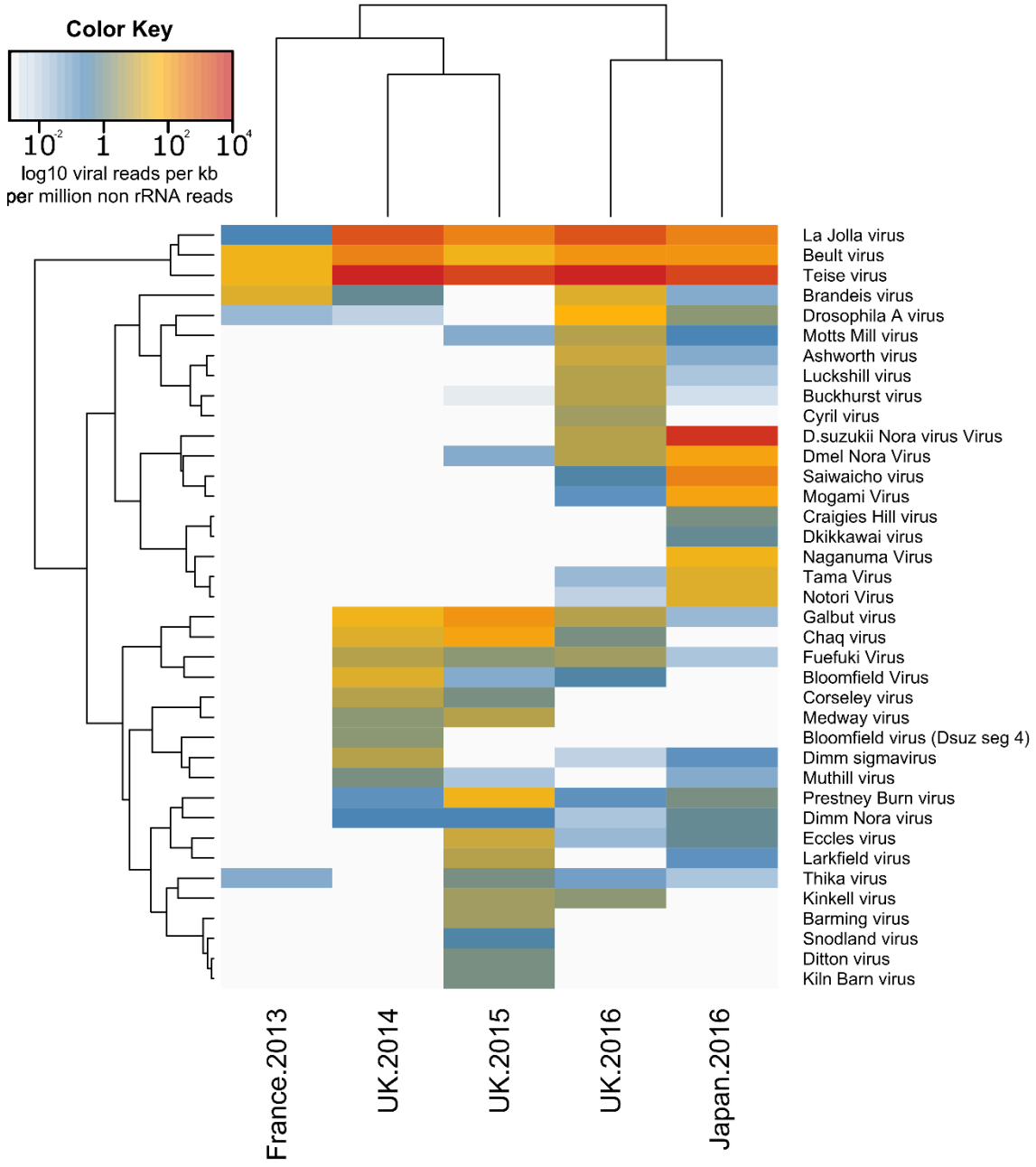


Figure 6. The heatmap shows the relative number of reads (log_{10} reads per kb per million non-ribosomal RNA reads) from each library mapping to each of the *Drosophila* viruses. Rows and columns are clustered by their similarity in read frequency on a log_{10} scale. A threshold for detection of 0.5 reads per kb per million non-rRNA reads was applied, however, a small amount of cross mapping is possible between closely related viruses and this may explain the detection of viruses with very low read counts. The low diversity of viruses in the France 2013 sample may be attributable to poly-A selection of RNA libraries. Some host species contamination and /or library cross-contamination (e.g. barcode switching) is possible in all samples, with a low proportion of COI reads mapping to *D. melanogaster* and *D. immigrans*. This could potentially explain the apparent presence of Sigma and Nora viruses from other drosophilids. Created using the 'heatmap2' function of the gplots package

bioRxiv preprint doi: https://doi.org/10.1101/190322; this version posted September 18, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supporting information

Figures

S1_Figure. Assembly of contigs of connecting two subgenomic fragments of Tiese virus. Motts Mill virus sequences (1 and 5) are here used here as references for the assembly. Sequences 2, 3 and 4 are contigs from UK *D. suzukii* reads, forming part of Tiese virus, one of which (3) had its closest BLAST hit to Prestney Burn virus and spans the gap between the two fragments of Tiese. PCR primer pair successfully spanning the fragments and producing amplicon of this 'bridge' segment are shown in S2_Table.

Tables

S1_Table. Known Drosophila Viruses detected in *D. suzukii* pools.

Viruses of Drosophila described by previous authors that were detected in our wild D. suzukii pools.

S2_Table. PCR Primer pairs for all newly discovered viruses

Primers used to detect new viruses by PCR. Melting temperature (T_m) as calculated by Primer3 plugin for Geneious 8.0.2. Annealing temperatures for used for reactions were universally 5°C lower than the stated melting temperature.

S3_Tables. Re-mapping of reads to known viruses

Number of reads mapping to all known drosophila viruses and to novel Dsuz viruses.

Data

S1_Data. Longest Contigs produced by Trinity

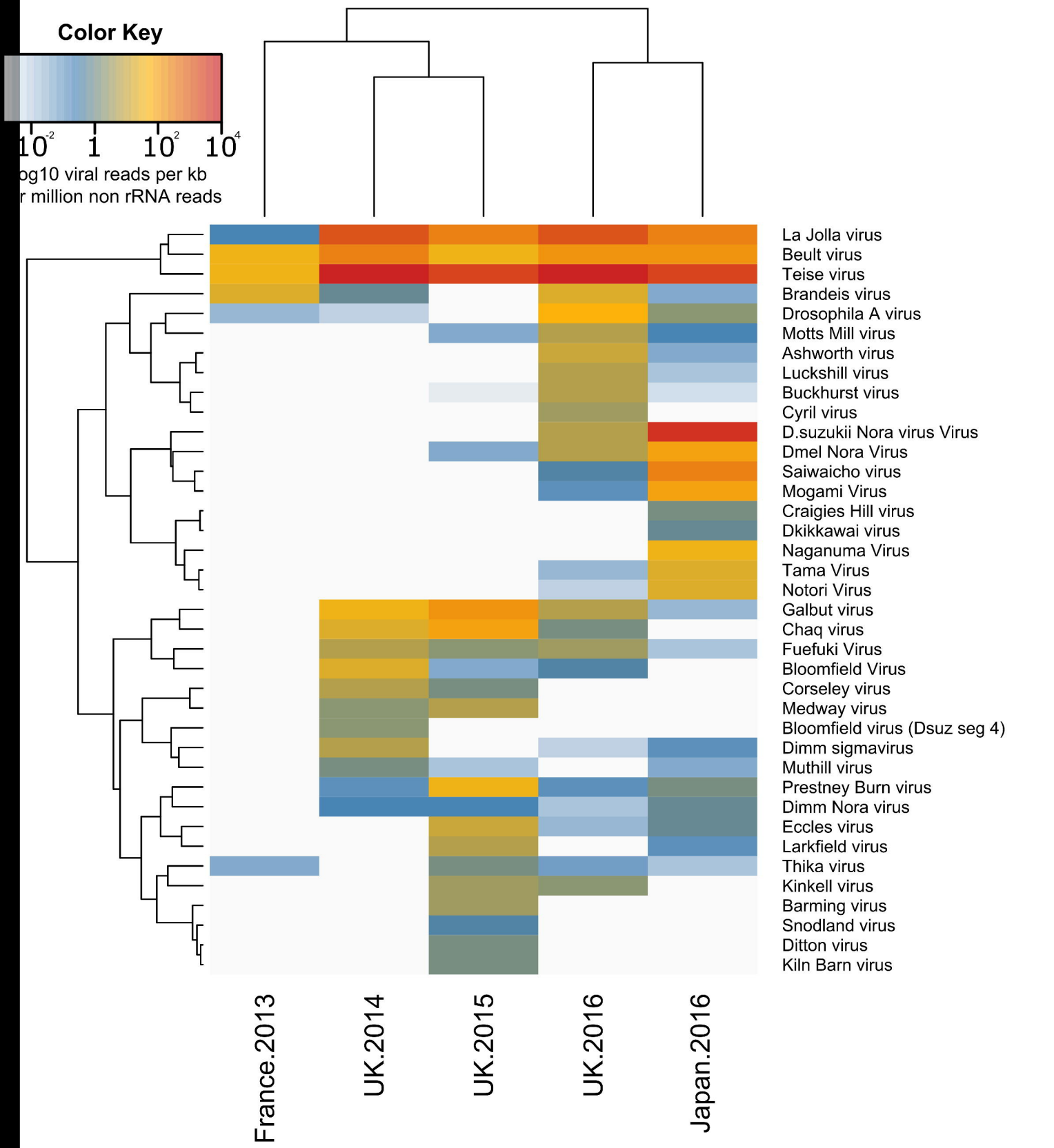
FASTA files containing all raw contigs produced by Trinity de novo assembly procedure (see Methods).

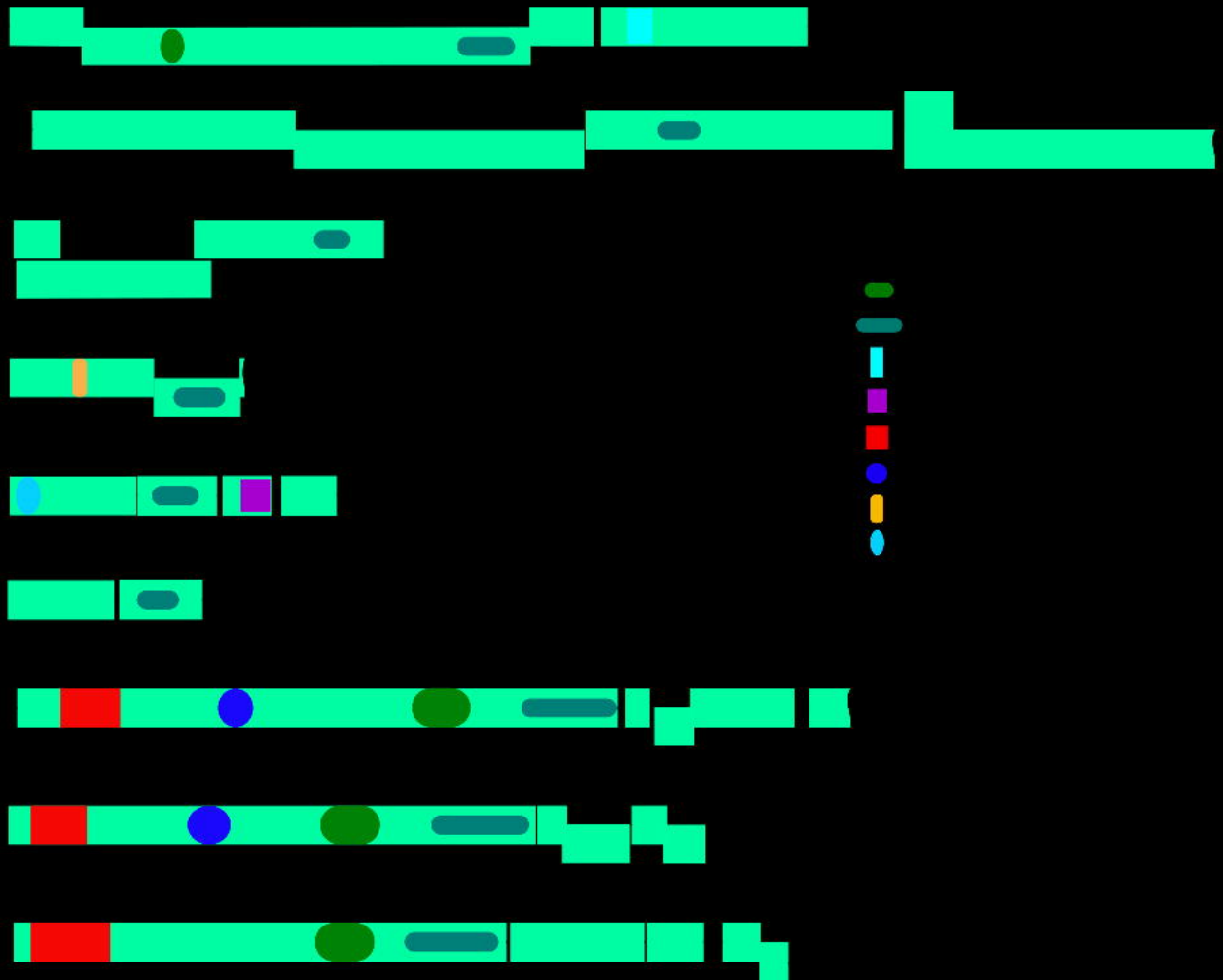
S2_Data Alignments of RdRp protein sequences used for phylogenetic analysis

Alignments of protein sequences aligned by ClustalW with Blossom cost matrix. New virus protein sequences are translations of contigs assembled from raw contigs produced by Trinity assembly neighbouring sequences are BLAST hits recovered from blastp searches of nr protein databases.

S3_Data Extended Maximum likelihood trees

Newick format, Maximum-likelihood trees inferred from viral polymerase sequences obtained by blastp.


Color Key
10^-2
1
10^2
10^4
og10 viral reads per kb
r million non rRNA reads
La Jolla virus
Beult virus
Teise virus
Brandeis virus
Drosophila A virus
Motts Mill virus
Ashworth virus
Luckshill virus
Buckhurst virus
Cyril virus
D.suzukii Nora virus Virus
Dmel Nora Virus
Saiwaicho virus
Mogami Virus
Craigies Hill virus
Dkikkawai virus
Naganuma Virus
Tama Virus
Notori Virus
Galbut virus
Chaq virus
Fuefuki Virus
Bloomfield Virus
Corseley virus
Medway virus
Bloomfield virus (Dsuz seg 4)
Dimm sigmavirus
Muthill virus
Prestney Burn virus
Dimm Nora virus
Eccles virus
Larkfield virus
Thika virus
Kinkell virus
Barming virus
Snodland virus
Ditton virus
Kiln Barn virus
France.2013
UK.2014
UK.2015
UK.2016
Japan.2016



Saiwaicho Virus

Medway virus

Tama virus

Beult Virus

Teise virus

Haemonchus contortus

Luckshill Virus

Fuefuki virus

Eutypa lata

Ceratitis capitata

Naganuma virus

Brandeis virus

Ceratitis capitata
Fopius arisanus
Bactrocera latifrons
Haematobia irritans

Drosophila suzukii Nora Virus

Colletotrichum cereale

Barming virus

Notori Virus

Procotyla fluviatilis

Aedes aegypti

Ditton virus translation

Mogami virus

Hyas araneus

Parasitella parasitica

Pseudocercospora fijiensis

Kiln Barn virus

Snodland virus

Melipona quadrifasciata
Lasius niger

Ceratitis capitata

Larkfield virus

Eccles virus