


# SCIENTIFIC REPORTS



OPEN

## Investigating the viral ecology of global bee communities with high-throughput metagenomics

David A. Galbraith<sup>1</sup> , Zachary L. Fuller<sup>1,2,3</sup>, Allyson M. Ray<sup>1</sup>, Axel Brockmann<sup>4</sup>, Maryann Frazier<sup>1</sup>, Mary W. Gikungu<sup>5</sup>, J. Francisco Iturralde Martinez<sup>6</sup>, Karen M. Kapheim<sup>7,8</sup>, Jeffrey T. Kerby<sup>9</sup>, Sarah D. Kocher<sup>10</sup>, Oleksiy Losyev<sup>11</sup>, Elliud Muli<sup>12</sup>, Harland M. Patch<sup>1</sup>, Cristina Rosa<sup>6</sup>, Joyce M. Sakamoto<sup>1</sup>, Scott Stanley<sup>13</sup>, Anthony D. Vaudo<sup>1</sup> & Christina M. Grozinger<sup>1</sup>

Bee viral ecology is a fascinating emerging area of research: viruses exert a range of effects on their hosts, exacerbate impacts of other environmental stressors, and, importantly, are readily shared across multiple bee species in a community. However, our understanding of bee viral communities is limited, as it is primarily derived from studies of North American and European *Apis mellifera* populations. Here, we examined viruses in populations of *A. mellifera* and 11 other bee species from 9 countries, across 4 continents and Oceania. We developed a novel pipeline to rapidly and inexpensively screen for bee viruses. This pipeline includes purification of encapsulated RNA/DNA viruses, sequence-independent amplification, high throughput sequencing, integrated assembly of contigs, and filtering to identify contigs specifically corresponding to viral sequences. We identified sequences for (+)ssRNA, (–) ssRNA, dsRNA, and ssDNA viruses. Overall, we found 127 contigs corresponding to novel viruses (i.e. previously not observed in bees), with 27 represented by >0.1% of the reads in a given sample, and 7 contained an RdRp or replicase sequence which could be used for robust phylogenetic analysis. This study provides a sequence-independent pipeline for viral metagenomics analysis, and greatly expands our understanding of the diversity of viruses found in bee communities.

Populations of bees around the world are exhibiting declines, which are the result of multiple interacting factors, including pressure from pathogens<sup>1</sup>. To date, more than 24 viruses have been identified in western honey bees (*Apis mellifera*)<sup>2–7</sup>. These viral infections can result in a range of symptoms, from no obvious phenotype to rapid death and colony loss, depending on the viral species, viral strain, physiological state of the host, and presence of other stressors (reviewed in McMenamin *et al.*<sup>8</sup>). However, despite the importance of bees as pollinators of flowering plants in agricultural and natural landscapes<sup>9,10</sup> and the importance of viruses to bee health, our understanding of bee viruses is surprisingly limited: despite the diversity of bee species and their worldwide distribution, the vast majority of the studies examining bee pathogens have focused on western honey bees (*Apis mellifera*) populations in North America and Europe<sup>11</sup>.

Bee viral ecology is particularly complex, since many viruses seem to be shared across diverse bee species (see review from Tehel, Brown & Paxton<sup>11</sup>). Several studies have demonstrated that viruses can spill-over from managed *A. mellifera* or bumble bee (*Bombus spp*) colonies to wild bee populations, with increased viral prevalence

<sup>1</sup>Department of Entomology, Center for Pollinator Research, Huck Institutes of the Life Sciences Pennsylvania State University, University Park, PA, USA. <sup>2</sup>Department of Biology, Pennsylvania State University, University Park, PA, USA. <sup>3</sup>Department of Biological Sciences, Columbia University, New York, NY, USA. <sup>4</sup>National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India. <sup>5</sup>Department of Zoology, National Museums of Kenya, Nairobi, Kenya. <sup>6</sup>Department of Plant Pathology, Pennsylvania State University, University Park, PA, USA. <sup>7</sup>Department of Biology, Utah State University, Logan, UT, USA. <sup>8</sup>Smithsonian Tropical Research Institute, Panama City, Panama. <sup>9</sup>Neukom Institute for Computational Science, Dartmouth College, Hanover, NH, USA. <sup>10</sup>Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. <sup>11</sup>Department of Beekeeping, The National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine. <sup>12</sup>The International Center of Insect Physiology and Ecology, Nairobi, Kenya. <sup>13</sup>Division of Pediatric Hematology and Oncology, Pennsylvania State University Children's Hospital, Hershey, PA, USA. Correspondence and requests for materials should be addressed to D.A.G. (email: [dag5031@gmail.com](mailto:dag5031@gmail.com))

in areas with increased density of infected managed colonies<sup>12–16</sup>. Viruses found to be pathogenic in *A. mellifera* (such as deformed wing virus, Israeli acute paralysis virus, acute bee paralysis virus, and Kashmir bee virus) have also been found to be pathogenic in wild bee species<sup>14,17,18</sup>. Targeted studies have identified subsets of the previously identified honey bee viruses (including deformed wing virus, black queen cell virus, sacbrood virus, Israeli acute paralysis virus, acute bee paralysis virus, Lake Sinai virus, chronic bee paralysis virus, and Kashmir bee virus) in bee populations in Asia<sup>19</sup>, Africa<sup>20</sup>, the Middle East<sup>21</sup>, Australia<sup>22</sup>, and South America<sup>23,24</sup>, suggesting a world-wide distribution of these viruses.

Since the most well-studied viruses were originally identified in *A. mellifera*, it has often been assumed that *A. mellifera* are the primary hosts and other bees are only secondarily infected. However, because many viruses seem to be so readily shared, it is not possible to determine which bee species, if any, serves as the primary host. Indeed, based on sequence analyses, the viral species that have been examined thus far seem to be shared within a geographic region rather than within a particular bee species, suggesting that there may not necessarily be bee species-specific viral strains<sup>13,14</sup>. Thus, it is likely that non-*Apis* populations also harbor viruses which can spill over into *A. mellifera*.

Similarly, recent metagenomics studies of bees have identified many viruses that were originally described in plants<sup>25,26</sup>. It has been assumed that these are simply contaminants from the pollen and nectar in the bees' diets, but these viruses may in fact be using bees as hosts. Indeed, tobacco ringspot virus (TRSV) was recently shown to infect and replicate in *A. mellifera*<sup>27</sup>. Further studies are needed to fully examine the intriguing possibility that pollinators are more than simply carriers of plant viruses.

Thus, fully understanding the viral community in managed and wild bee species, and the plants they feed on, requires a holistic approach that examines these communities in diverse geographic regions. Distinct types and species of viruses undoubtedly remain unidentified in populations of *A. mellifera* and other bee species across the world. Moreover, due to increased globalization, it is likely only a matter of time before these viruses (and other pathogens and parasites) are spread to new regions. For example, recently a colony of an exotic stingless bee species (*Plebeia frontalis*) was found to be established in California<sup>28</sup>, potentially harboring pathogens that could be spread to other bee species within the region.

With the development and optimization of inexpensive high-throughput sequencing technologies, it is now possible to sequence the genetic material of the entire community of viruses present within a host, without any prior knowledge of the viral genome sequences (viral metagenomics analysis)<sup>29</sup>. This technology allows rapid and efficient identification of previously uncharacterized types and strains of viruses, bypassing labor-intensive steps of purifying and culturing the viruses. Viral metagenomics has been utilized by many studies seeking to identify novel viruses and characterize viral communities in diverse organisms (reviewed in Mokili *et al.*<sup>30</sup>).

Using this approach, several recent studies have identified novel viruses in populations of bees (*A. mellifera*, *Bombus terrestris*, *Bombus pascuorum*, *Osmia cornuta*, and *Andrena vaga*) collected from sites in the Netherlands, South Africa, Tonga, Israel and Belgium<sup>6,26,31</sup>. However, while these studies have provided valuable information about viral communities within particular species or regions, they are limited in scope, in that they focus on a single bee species in multiple locations (*A. mellifera* populations in Israel<sup>31</sup> or the Netherlands, South Africa, and Tonga<sup>6</sup>) or a selection of wild bees in a single region (Belgium<sup>26</sup>).

To improve our ability to detect viruses, we have developed a protocol for unbiased, sequence-independent detection of novel viruses within diverse bee populations from around the globe. Additionally, we used a viral enrichment protocol to purify the viral particles with the aim of increasing the proportion of sequence reads obtained from viruses versus the host bee species, compared to previous studies that did not use this protocol<sup>6,26</sup>. We used this approach to characterize the viral communities of 12 bee species collected from 9 countries across 4 different continents and Oceania (see Table 1 for bee species and their respective locations). This approach allowed us to search for novel viruses, begin to determine the distribution of these viruses (and viral families) across the world, and begin to evaluate the extent to which *A. mellifera* and wild bee populations share these viruses. As in other studies<sup>25,26</sup>, we identified several viruses that share sequence homology to plant viruses and thus may be the result of contamination, but given that these were found in multiple populations and locations, it raises the question of whether or not these viruses may use bees as hosts.

## Methods

**Sample collection.** Sample collection kits and standardized protocols were provided to collaborators to obtain samples of foraging bees from geographically distant locations. For each geographic region, collaborators were instructed to collect western honey bees (*Apis mellifera*) and other common bee species (including both *Apis* and non-*Apis* bee species) as individuals foraged on flowering plants; no attempt was made to preferentially collect individuals that appeared diseased. Individuals from different bee species were collected in separate tubes to ensure that the species were not mixed. At least 20 individuals were collected onto ice-cold 95% ethanol for each species from each region. Details for each site were recorded, including the GPS coordinates. A summary of different locations and sampled species can be found in Table 1. The samples were then shipped to Pennsylvania State University for processing. We also included samples of *A. mellifera* with deformed wings (a symptom of high levels of DWV) collected from a colony at the Penn State University apiaries as a control for DWV detection (labeled Positive Control in Table 1).

**Virus purification and nucleic acid extraction.** 10 individuals from each sample group (representing a specific species collected at a specific location) were placed in individual 2.0 ml microcentrifuge tubes with 150  $\mu$ l nuclease free H<sub>2</sub>O and 3–5 sterile glass beads and homogenized in a FastPrep FP120 cell disruptor (Thermo-Fisher Scientific, Waltham, MA, USA) for 45s. 100  $\mu$ l of the homogenate from each of the 10 individuals was pooled in a common 1.5 ml microcentrifuge tube. Due to their large size, 10 *Xylocopa* spp. samples were pooled and homogenized using a mortar and pestle in 5 ml nuclease free H<sub>2</sub>O. The pooled homogenate from

Sample_ID	Location	Species	Original ID	GPS Coordinates (°Lat., °Long.)
Positive Control	Pennsylvania, USA (DWV control sample)	<i>Apis mellifera</i>	DWV	40.8198, -77.8588
PA1-Am	Pennsylvania, USA	<i>Apis mellifera</i>	C21-S2	40.8369, -77.8786
PA2-Am	Pennsylvania, USA	<i>Apis mellifera</i>	C22-S3	40.7106, -77.9645
PA3-Am	Pennsylvania, USA	<i>Apis mellifera</i>	C23-S4	40.7534, -77.6699
PA4-Am	Pennsylvania, USA	<i>Apis mellifera</i>	SCAm	40.8198, -77.8588
PA-Bi	Pennsylvania, USA	<i>Bombus impatiens</i>	SCBi	40.8198, -77.8588
WA-Mr	Washington, USA	<i>Megachile rotundata</i>	WSmR	46.0177, -118.6975
WA-Nm	Washington, USA	<i>Nomia melanderi</i>	WSNm	46.0177, -118.6975
WA-Am	Washington, USA	<i>Apis mellifera</i>	WSAm	46.0177, -118.6975
CA-Am	California, USA	<i>Apis mellifera</i>	C24-C1	41.7928, -124.1208
NI-Am	Nicaragua	<i>Apis mellifera</i>	NIAm	12.1184, -86.1224
NI-Ts	Nicaragua	<i>Trigona silvestriana</i>	NIST	12.1184, -86.1224
PN-Am	Panama	<i>Apis mellifera</i>	PNAm	9.1500, -79.8500
PN-Mc	Panama	<i>Megalopta centralis</i>	PNMg	9.1500, -79.8500
GE-Am	Georgia	<i>Apis mellifera</i>	GAAm	42.0681, 42.4403
CH-Am	Switzerland	<i>Apis mellifera</i>	SWAm	46.5197, 6.6323
CH-Bs	Switzerland	<i>Bombus spp.</i>	SWNb	46.5197, 6.6323
UA-Am	Ukraine	<i>Apis mellifera</i>	UKAm	50.4501, 30.5234
KE1-Am	Kakamega, Kenya	<i>Apis mellifera</i>	C11-G6	0.2527, 34.9232
KE2-Am	Kakamega, Kenya	<i>Apis mellifera</i>	C6-G1	0.2277, 34.8459
KE3-Am	Kisii, Kenya	<i>Apis mellifera</i>	C4-K1	-0.7423, 34.8045
KE4-Am	Kisii, Kenya	<i>Apis mellifera</i>	C5-K2	-0.7399, 34.8895
KE5-Am	Mbita, Kenya	<i>Apis mellifera</i>	C16-M2	-0.4803, 34.2929
KE6-Am	Mbita, Kenya	<i>Apis mellifera</i>	KNAm	-0.4354, 34.2144
KE7-Am	Naivasha, Kenya	<i>Apis mellifera</i>	C1-N1	-0.07547, 36.4384
KE8-Am	Naivasha, Kenya	<i>Apis mellifera</i>	C2-N2	-0.5713, 36.3603
KE9-Am	Naivasha, Kenya	<i>Apis mellifera</i>	C3-N3	-0.6616, 36.3835
KE-XY	Mbita, Kenya	<i>Xylocopa spp.</i>	KNXY	-0.4734, 34.1823
IN-Ac	UAS-GKVK, Bangalore, India	<i>Apis cerana</i>	INAc	13.0774, 77.5778
IN-Ad	UAS-GKVK, Bangalore, India	<i>Apis dorsata</i>	INAd	13.0774, 77.5779
IN-Af	UAS-GKVK, Bangalore, India	<i>Apis florea</i>	INAf	13.0711, 77.5802
IN-Am	UAS-GKVK, Bangalore, India	<i>Apis mellifera</i>	INAm	13.0712, 77.5805
IN-Ti	UAS-GKVK, Bangalore, India	<i>Tetragonula iridipennis</i>	INST	13.0752, 77.5800
NZ1-Am	Hamilton, New Zealand	<i>Apis mellifera</i>	NZN1	-37.7870, 175.2793
NZ2-Am	Tauranga, New Zealand	<i>Apis mellifera</i>	NZN2	-37.6878, 176.1651
NZ3-Am	Christchurch, New Zealand	<i>Apis mellifera</i>	NZS1	-43.5321, 172.6362
NZ4-Am	Otago, New Zealand	<i>Apis mellifera</i>	NZS2	-44.8280, 169.6345

**Table 1.** Sample information, including ID, collection location, species, and GPS coordinates.

each sample was then dialyzed using a 0.2 µm cell filter (Corning, Tewksbury, MA, USA) to purify the virus extracts as in Hunter *et al.*<sup>32</sup> and Liu *et al.*<sup>33</sup>. A 125 µl aliquot of each sample was treated with a nuclease cocktail (including 14U Turbo DNase I, 25U Benzonase, 20U RNaseI, and 10X DNase buffer) and incubated at 37 °C for 1.5 hours to remove all nucleic acid that was not protected by a viral capsid as in He *et al.*<sup>34</sup> Nucleic acids (both RNA and DNA) from the purified encapsulated viruses were extracted using a MagMAX Viral Isolation Kit (Thermo-Fisher Scientific, Waltham, MA, USA), according to the manufacturer's protocol.

**Sequence independent nucleic acid amplification and sequencing.** A previously established protocol was used to obtain unbiased random amplification of the extracted nucleic acids<sup>35</sup>. Briefly, the first strand cDNA was synthesized from 30 µl of viral nucleic acid using a random primer design (consisting of a 20 base oligonucleotide sequence followed by a randomized octamer sequence: GACCATCTAGCGACCTCCACNNNNNNNN) in a reverse transcriptase (RT) reaction using a High-Capacity cDNA Reverse Transcription Kit (Thermo-Fisher Scientific, Waltham, MA, USA). To synthesize the second strand, the 19 µl of the initial cDNA was denatured at 95 °C for two minutes and cooled to 4 °C prior to the addition of Klenow fragment DNA polymerase (New England Biolabs, Ipswich, MA, USA) for fragment extension at 37 °C for 60 minutes. Finally, PCR amplification was performed for 40 cycles using the above primer without the randomized octamer sequence (GACCATCTAGCGACCTCCAC) and 2 µl of dsDNA template. The PCR reactions were then purified using a MSB Spin PCRapace Purification Kit (Invitex, Berlin, Germany) using the manufacturer's standard protocol.

Quality and concentration of the samples was assessed using a 2100 Bioanalyzer (Agilent, Santa Clara, CA). The samples were then submitted to the Genome Core Facility at Penn State University for sequencing on the Illumina MiSeq, resulting in 1,056,165 (+/− 105,794) reads at 150 bases per sample (see Supplementary Table 1 for more information on the number of reads and assembled contigs in each sample). Although the phenomenon known as “index switching” has been observed to misattribute reads to the wrong sample in multiplexed sequencing libraries in the same run on certain Illumina platforms, this effect is demonstrated to be limited on the MiSeq and is not expected to generate increases in external contaminants<sup>36</sup>. Raw data from this Whole Genome Shotgun project have been deposited at DDBJ/ENA/GenBank under the accession SAMN07695258, along with BioProject (PRJNA407112) BioSamples (SAMN07634920–SAMN07634956).

**Data processing.** The quality of raw sequencing reads was first manually inspected using FastQC<sup>37</sup>. Illumina adaptor sequences and reads with low quality scores (25) were removed using Trimmomatic v0.30<sup>38</sup>. As the randomized octomer sequences correspond to regions of viral genome and not every read contains them, we did not trim the start of each read, although we note that polymorphisms in these regions may possibly complicate the assembly. To identify novel viruses, the trimmed and filtered sequencing reads were assembled *de novo* using three different programs: Oases v0.2.8/velvet v1.2.10 (N50: 310 nt)<sup>39</sup>, Trinity v2.1.0 (N50: 326 nt)<sup>40</sup>, and SPAdes v3.7.1 (N50: 315 nt)<sup>41</sup>. The N50 (weighted median length of the contigs) for all assemblies were low (the total read length was 150 nt), so the three assemblies for each sample were merged using the software package Minimus2 (AMOS v3.1.0)<sup>42</sup>, resulting in an improved N50 of 699 nt. For comparison, this N50 is greater than for similar previous studies of *A. mellifera* and *Culex pipiens* (mosquito) associated viruses<sup>6,43</sup>, yet we do not exclude the possibility that further optimization of Kmer sizes could yield slight increases in overall contig length. For the three assembly methods, we used Kmer sizes of 21 (Oases/velvet) and 25 (Trinity), while SPAdes uses multiple kmer sizes (21, 33, and 55) to identify the optimal assembly.

The assembled contigs were then subjected to a series of processing steps. To remove contamination and retain a set of candidate contigs likely to be viral in origin, each assembly was screened for homologous sequences from three potential sources of contamination: the human genome (version hg38), all bacterial genomes in the NCBI database, and the honey bee genome (version amel4.5) using an E-value cutoff of  $10^{-5}$ . The total set of contigs was screened for sequence homology to the three possible sources of contamination using BLAST<sup>44</sup>. Contigs with significant hits to any of the contamination sources were then removed from further analyses (see Supplementary Table 1 for the number of contigs matching each source of contamination). To identify the closest known viral sequence homology, the remaining assembled contigs were compared to all sequences in the NCBI nr database<sup>45</sup> using DIAMOND v0.8.36<sup>46</sup> with an E-value cutoff of  $10^{-5}$ . Viral families were assigned using the 2017 report of the International Committee on Taxonomy of Viruses (ICTV) for the classification and nomenclature of viruses<sup>47</sup> (see Supplementary Table 2).

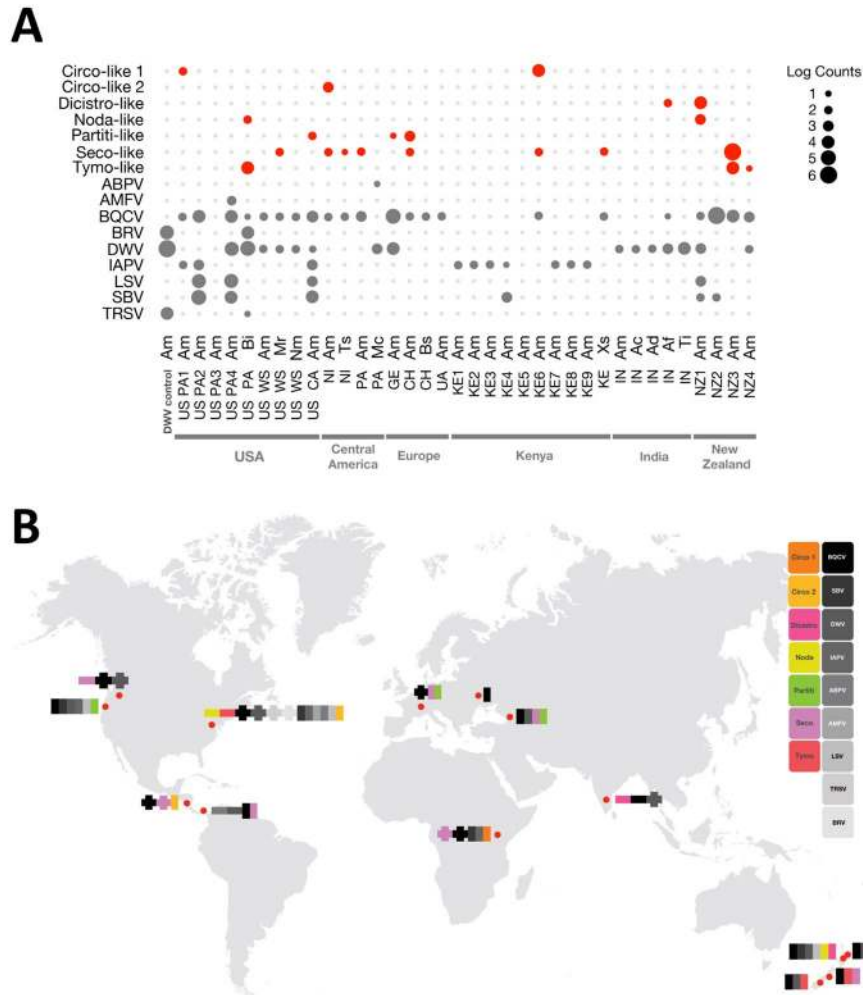
To determine relative abundances of each virus, the trimmed sequence reads from each sample were then mapped to the set of candidate viral contigs identified in that sample using Tophat2 v2.0.9<sup>48</sup> (see Supplementary Table 2). Contigs with >0.1% of the total reads successfully mapping in that sample were chosen for further analyses (see Supplementary Table 3). For the contigs with >0.1% of the total reads, ORFfinder<sup>49</sup> was used to identify potential open reading frames within the contigs that had significant BLAST hits (E-value <  $1 \times 10^{-5}$ ) to known viruses to further characterize the viral contigs from each sample. Viral contigs that showed evidence of a putative RNA dependent RNA polymerase (RdRp) or other replicase proteins were selected for further phylogenetic analyses. Note that while we used a sequence independent amplification approach to generate the libraries for sequencing, amplification bias may still have skewed the total numbers of reads for the viral sequences.

**Phylogenetic analysis.** Reference sequences corresponding to the RdRp or other replicase proteins were downloaded from NCBI GenBank<sup>50</sup> for viruses that were closely related to the significant BLAST hit to identify the relationship of the novel viruses discovered here to other known viruses. Multiple amino acid sequence alignments were performed using MUSCLE<sup>51</sup> in MEGA7<sup>52</sup> with default settings. Phylogenetic trees for each novel virus identified in Fig. 1 were created using the maximum-likelihood method with 500 bootstrap replications in MEGA7<sup>52</sup>, by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with the superior log likelihood value.

## Results

**Identification of Viral Sequences.** On average, 6.87% of assembled contigs in each sample had evidence of being viral in origin (see Supplementary Table 1). Our enrichment protocol is aimed to maximize the amount of viral nucleic acids in each sample, which are typically far fewer than found in host or other microbes<sup>53</sup>. Indeed, in previous studies of viruses found in bee species, the total proportion of reads determined to originate from viruses range from 0.1 to 2.7%<sup>6,26</sup>. However, we recognize further improvements are needed to increase the abundance of viral genomic material in metagenomics studies of bees, and note that in smaller insect species such as *Ixodes scapularis* ticks the total proportion of viral reads has been reported to be as high as 35%<sup>54</sup>.

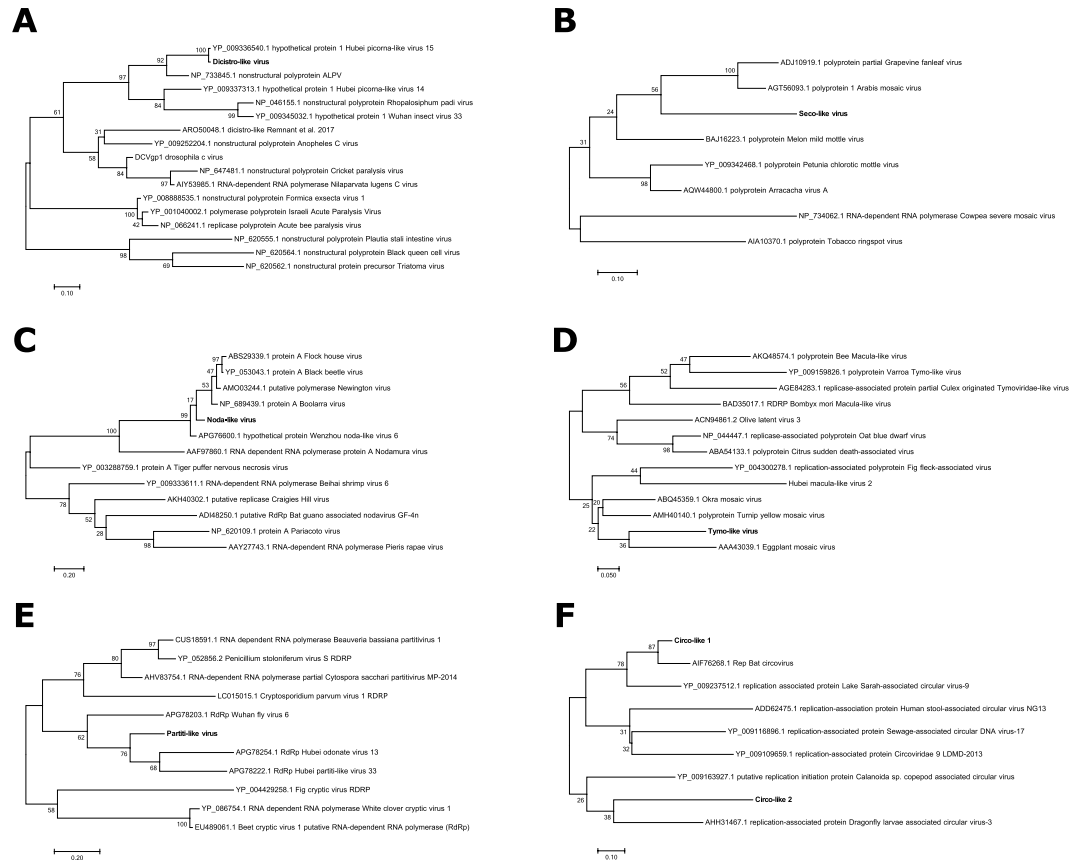
With our approach, we identified 479 different contigs among the 37 samples that shared sequence homology with previously identified viral sequences (see Supplementary Table 2 for information on each of the contigs identified in each sample). Of these, 308 contigs correspond to “novel” viruses (defined as viruses that were not previously found in bees), and 223 contigs correspond to known bee viruses. Many of these contigs were identified in multiple locations and/or species. When duplicate/overlapping contigs are removed, the number of unique novel virus contigs was reduced to 127. Next, we examined the number of contigs that were supported by >0.1% of the total reads for the given sample in which it was found, as these may represent more prevalent or abundant viruses in bee populations. Only 27 contigs matched these criteria (see Supplementary Table 3).



**Figure 1.** Summary of known and newly identified viral contigs from all samples. These figures summarize the distributions of the known viruses and newly identified viruses (with RdRp sequences and present at >0.1% of the reads in a given sample) from this study. **(A)** Plot showing the prevalence of all known bee viruses (in grey), as well as the 7 novel viruses detected in this dataset that contain an RdRp or other replicase protein (in red) in log read counts. Individual samples are listed according to location, with each sample name containing the location (two letter code for country or, for US samples, state) followed by a two letter code for the species. See Table 1 for complete information on location and species for each sample. **(B)** This map shows the distribution of known viruses (colored in greyscale) and newly identified viruses (colored) in *A. mellifera* samples (vertical bars) and non-*A. mellifera* samples (horizontal bars) collected at each site. Where a virus was identified in both sample types, the horizontal and vertical bars are overlapping.

Sequence homology analyses of these 27 novel virus contigs revealed 9 different viral families among these bee associated viruses, including 6 viral families that had not been heretofore observed in bees (noted in Supplementary Table 3). Of these, one virus from the family *Rhaboviridae* (BRV) was recently described in *A. mellifera* populations in the United States, Israel, South Africa, Tonga and the Netherlands, and *B. impatiens* populations in the United States<sup>6,55</sup>. Of the 27 novel virus contigs, 7 were found to contain a putative RdRp or replicase that could be used for phylogenetic analyses.

The distribution and relative quantities of the known viruses and newly identified viruses with RdRp domains (that were present at >0.1% of the reads in each sample) across each sample are shown in Fig. 1A,B provides information on the geographic distribution of these viruses and their presence in *A. mellifera* samples versus non-*A. mellifera* samples at each site. Details about each of these novel viruses, including their phylogenetic characterization, are presented below. Interestingly, 5 novel viruses were identified in North American samples, 4 in New Zealand, and 2 in each of the other continents; thus, though the North American samples are relatively well-studied compared to the other regions, at least in our analysis, these samples still generated more novel viruses. In regions where both *A. mellifera* and non-*A. mellifera* bees were sampled, *A. mellifera* had higher numbers of viruses in four regions (Kenya, Nicaragua, Switzerland, and Pennsylvania, US), lower numbers in two regions (India and Washington, US) and equivalent numbers in Panama. Only in the United States and India were



**Figure 2.** Phylogenetic analyses of novel virus contigs. Maximum-likelihood phylogenetic trees showing the evolutionary relationships between the novel virus contigs containing an RdRp or other replicase protein and other representative sequences from (A) *Dicistroviridae*, (B) *Secoviridae*, (C) *Nodaviridae*, (D) *Tymoviridae*, (E) *Partitiviridae*, and (F) *Circoviridae*. NCBI Accession numbers for the sequences used in these trees can be found in the figure above.

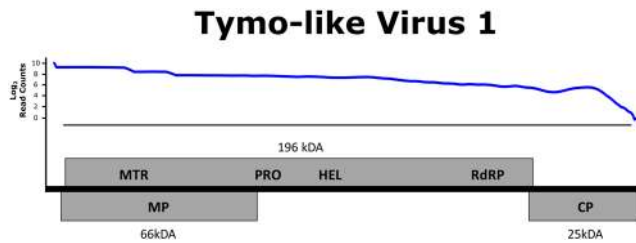
novel viruses found in non-*A. mellifera* samples that were not present in the *A. mellifera* samples as well. These results suggest that these viruses are globally distributed, both across bee species and geographic regions.

Below, we discuss the results for the known viruses, as well as the 7 viruses that possess a putative RdRp or replicase protein.

**Known Viruses.** Within this dataset, we identified 7 well-studied bee viruses: black queen cell virus (BQCV), deformed wing virus (DWV), Israeli acute paralysis virus (IAPV), Lake Sinai virus (LSV), acute bee paralysis virus (ABPV), sacbrood virus (SBV), and *Apis mellifera* filamentous virus (AMFV). BQCV represented the most common known virus among all of the samples, with 22 of the 36 samples (DWV positive control sample excluded) testing positive (61%). DWV, another common virus, was detected in 16 (44%) of the samples, and was the second most prevalent virus detected in this study. IAPV (10 samples, 28%), SBV and LSV (6 samples, 17%), and ABPV and AMFV (1 sample, 3%) conclude the known bee viruses identified in our samples (see Fig. 1 and Supplementary Table 2 for more details on the locations and species in which these viruses were found).

**Newly Identified Positive Sense RNA Viruses.** *Dicistro-like virus.* We identified one *Dicistro-like* contig that shared sequence homology with known viruses from the family *Dicistroviridae*, a viral family that includes several known bee-infecting viruses, such as IAPV, BQCV, and Aphid lethal paralysis virus (ALPV). This virus was found in *Apis florea* samples from India, with a contig length of 3,818 nt, which included a putative RdRp region, as well as in *A. mellifera* samples from New Zealand with a length of 1,382 nt. A phylogenetic analysis of this region revealed that this particular contig was most closely related to aphid lethal paralysis virus (ALPV), and forms a clade with ALPV and other dicistroviruses identified in a previous study using other invertebrates, including odonates, heteropterans, myriapods, and orthopterans (Fig. 2A)<sup>56</sup>.

*Seco-like virus.* We identified a *Seco-like* virus that shares sequence homology to *Secoviridae*, which is in the same viral order as *Dicistroviridae* (Picornavirales). Overlapping viral contigs were found in four bee species across 4 continents and Oceania (*A. mellifera* in Georgia: 1,988 nt, Kenya: 971 nt, Panama: 1,295 nt, Nicaragua: 1,955 nt, Switzerland: 1,105 nt, and New Zealand: 1,799, *Megachile rotundata* in the US (Washington): 1,500 nt,



**Figure 3.** Genome of the newly discovered Tymo-like virus. Genome structure of the Tymo-like virus discovered in *Bombus impatiens* and *Apis mellifera* from the United States and New Zealand. This viral contig contains a long polyprotein encoding a methyltransferase (MTR), proteinase (PRO), helicase (HEL), and an RNA dependent RNA polymerase (RdRP) that is common among viruses from the Tymovirus genus. Additionally, a separate overlapping protein encoding a movement protein at the 5' end and coat proteins was detected at the 3' end of the viral contig. Log read counts were calculated using a sliding window of 100 bases.

*Trigona silvestriana* in Nicaragua: 950 nt and *Xylocopa spp.* from Kenya: 1,254 nt). There were minimal differences across the sequences: the putative RdRp region was >99% similar across the samples. A phylogenetic analysis of the RdRp from this virus suggests close sequence homology to a group of plant viruses (including Arabis mosaic virus; see Fig. 2B).

**Noda-like virus.** We identified a novel (+)ssRNA virus within the family *Nodaviridae* in two different species, *A. mellifera* (1,216 nt) and *B. impatiens* (1,831 nt). Interestingly, these were from two distant geographic locations (Hamilton, New Zealand and Pennsylvania, United States). Using a BLASTx<sup>44</sup> approach on the open reading frames identified within the assembled contigs, we identified a putative RdRp that was used for phylogenetic analysis. The phylogenetic analysis revealed that this virus was most closely related to other viruses within the *Nodaviridae* family of viruses, forming a clade with members of the viral genus *Alphanodavirus* (Fig. 2C), which includes several insect viruses, including flock house virus, black beetle virus, and nodamura virus.

**Tymo-like virus.** A Tymo-like virus with overlapping contigs was identified in *A. mellifera* from two different locations in New Zealand (Christchurch: 6,130 nt and Otago: 276 nt) and *B. impatiens* from Pennsylvania, United States (1,670 nt), sharing sequence homology with the viral family *Tymoviridae*. This viral family is comprised predominantly of plant viruses, however, a recent study also identified a virus from this family in *A. mellifera* samples<sup>5</sup>. The novel virus described here exhibits similar genomic characteristics to other viruses in the Tymovirus genus, with a long poly protein that encodes a methyltransferase (MTR), proteinase (PRO), helicase (HEL), and an RNA dependent RNA polymerase (RdRP), a separate overlapping protein that encodes a movement protein, and a coat protein at the 3' end (see Figs 2D and 3).

**Newly Identified Negative Sense RNA Viruses.** Only recently have negative sense RNA ((-)ssRNA) viruses been identified in bees<sup>6,26</sup>. One of these, *Apis mellifera* Rhabdovirus 1 (ARV-1), from the family *Rhabdoviridae*<sup>6</sup>, was detected in *A. mellifera* populations in South Africa, Tonga, and the Netherlands. It was also recently identified in bee populations in Israel and the United States<sup>55</sup>. Our data sets related to this virus were published elsewhere<sup>55</sup> and thus we will only briefly summarize this information here.

We identified contigs corresponding to this Rhabdovirus in two different species (*A. mellifera* and *Bombus impatiens*) that were collected in Pennsylvania, United States, suggesting this virus can easily switch hosts. Thus, Levin *et al.*<sup>55</sup> proposed a different nomenclature, Bee Rhabdovirus (BRV). The two assembled contigs for each species were 13,844 nt and 14,589 nt in *A. mellifera* and *B. impatiens* respectively, and were >99% similar to ARV-1. These assembled contigs contained the 5 genes that are typically carried by Rhabdoviruses, including the large protein (L), a glycoprotein (G), nucleoprotein (N), phosphoprotein (P), and matrix protein (M). Further information about this virus can be found in Levin *et al.*<sup>55</sup>.

**Newly Identified Double Stranded RNA Viruses.** *Partiti-like virus.* Overlapping contigs with sequence homology to other viruses within the viral family *Partitiviridae* were identified in *A. mellifera* samples obtained from California, United States (1,139 nt), Switzerland (409 nt), and Georgia (385 nt). A phylogenetic analysis of a putative RdRp region from the assembled contigs, determined by a BLASTx<sup>44</sup> screen, revealed sequence homology that is similar to other identified partitiviruses, but most closely related to the newly identified arthropod partitiviruses<sup>56</sup> (Fig. 2E).

**Newly Identified Single Stranded DNA viruses.** *Circo-like virus.* Sequences corresponding to a circo-like virus in *A. mellifera* from three different locations (Pennsylvania, United States: 521 nt, Kenya: 688 nt, and Nicaragua: 1,531 nt) with sequence homology to viruses from the *Circoviridae* family. Interestingly, a phylogenetic analysis of the replicase protein revealed two different viruses, with the sequences from the US and Nicaraguan *A. mellifera* populations being closely related to circular ssDNA viruses found in bats and dragonflies, respectively (Fig. 2F). The Circo-like virus contig that was discovered in Kenya shared sequence homology with bat circovirus (similar to the virus in the Pennsylvania samples), but did not overlap with the replicase protein.

**Confirmation of viruses using PCR.** We used PCR to confirm the presence of the expected viral contigs in the samples for BRV (see Levin *et al.*<sup>55</sup>), Circo-like 2, Noda-like, Tymo-like, Seco-like, and Dicistro-like viruses (see Supplementary Materials for primers, amplicon size, bee sample set).

## Discussion

In this study, we developed a viral metagenomics pipeline to comprehensively examine the viromes of specimens of 12 different bee species collected from a broad array of locations (9 countries, across 4 continents and Oceania). This dataset dramatically enhances our current knowledge of bee associated viruses, by identifying 27 novel (heretofore unidentified in bees) viral contigs from 9 viral families that were present at >0.1% of the reads within a sample. Six of these contigs contained a putative RdRp or replicate protein region (identified by BLASTx<sup>44</sup> and ORFfinder<sup>49</sup>) that could be used for phylogenetic analyses to determine relationships between previously identified viruses (see Fig. 2). Moreover, using our pipeline, we identified a wide variety of new types of viruses that had not been previously described (or only recently described) in bees, including (–)ssRNA, dsRNA, and ssDNA viruses.

This study represents the largest effort to identify novel pathogens in global bee samples to date. However, it is important to note that this study still had relatively small sample sizes, given the diversity of bee species and their global distributions. Sampling 10 bees per population (as in our study) is sufficient to detect more prevalent viruses within a given area, however, this sampling effort may be limited in the detection of low-prevalence viruses<sup>57</sup>. Indeed, by focusing on viruses that were represented in >0.1% of the reads in a sample, we may have further removed data from low abundance viruses. We were also unable to assemble full length genomes for most of the viruses detected, which may be due to a number of factors, including low sequencing coverage, uneven sequence coverage across viral genomes, and low sequence homology to sequences in current databases. Moreover, while our approach to enrich our samples for encapsulated viruses may have improved our ability to detect viruses by reducing the reads coming from host material (indeed, in 30/37 samples, more contigs matched to the viral databases than the honey bee genome, see Supplementary Table 1), this method involves additional processing steps which may have degraded some types of viruses. Furthermore, our conservative bioinformatics filtering procedure (E-value <  $1 \times 10^{-5}$ ) retained only a set of contigs with highly significant matches to viruses described in the NCBI nr database, thereby potentially removing sequences with homology to uncharacterized viral families (in only 3/37 samples, more contigs matched the viral databases than all potential sources of contamination and unmapped sequences combined). To limit the amount of genomic material generated from microbial contaminants, it may also be possible to avoid using midgut tissue which likely harbors a greater number of bacteria than other tissues. Thus, additional screens using larger sample sizes, both pooled and individual samples, alternative extraction techniques, and greater sequencing depth may help improve the recovery of larger contigs, potentially identifying additional viruses and likely demonstrating a broader geographic and species range for the viruses we identified in this study.

Additionally, the goal of this study was to identify viruses that may be circulating among co-foraging bee populations. Thus, we exclusively sampled foraging bees in the landscapes in the summer months. Viral titers can vary greatly among bees in different developmental stages, behavioral states, or seasons<sup>8</sup>. Moreover, highly diseased bees exhibit reduced foraging<sup>58</sup>. Additional studies using a broader diversity of bee samples – including samples of bees exhibiting clear symptoms of disease – may result in a broader array of identified viruses.

It is important to note that we did not assess replication of these viruses in our samples, and this will be necessary to confirm that bees are indeed hosts of these viruses. Viruses may be present in bee samples due to contamination from the environment or their diet; indeed, viruses can infect or contaminate pollen grains and be transmitted to individuals foraging on or consuming this pollen<sup>12,59</sup>. We selected viruses that were present at >0.1% of the reads within a sample (an average of ~1,000 reads per sample), to reduce the likelihood that we were examining a trace contaminant. Additionally, all of the viruses were identified in multiple species and multiple sites in independent collections and extractions, and thus it seems somewhat unlikely that these viruses were simply contaminants. Moreover, recent screens of samples collected as part of the US National Honey Bee Disease Survey<sup>60</sup> have identified several of these viruses in samples from across the country (Ray *et al.*, in preparation).

Positive sense RNA ((+)ssRNA) viruses are the most common type of virus that have been identified in bees thus far, with a vast majority of the viruses possessing this genome architecture. Most of these viruses belong to the viral order Picornavirales (including DWV, BQCV, IAPV, and SBV), but recently additional studies have added to the list of viral orders that infect bees, including Tymovirales<sup>6,26</sup>. Here we have identified novel virus sequences that share sequence homology with three (+)ssRNA virus families, *Dicistroviridae*, *Secoviridae*, and *Tymoviridae*.

Dicistroviruses are commonly found in honey bee colonies, with the most common viruses being IAPV and BQCV<sup>61</sup>. Unsurprisingly, we found BQCV to be the most widely distributed virus in our dataset, while IAPV was also fairly common, following only behind BQCV and DWV. Another recent study also identified a novel dicistrovirus in *A. mellifera* from Amsterdam<sup>6</sup>, however this virus is phylogenetically different than the virus we described here. We found this viral contig (containing a putative RdRp) in *A. florea* from India, however, we also detected another viral contig from *A. mellifera* from New Zealand that shared sequence homology with a different region of the same virus (Hubei picorna-like virus 15; accession number: NC\_032757.1). Unfortunately, there was no overlap between these two contigs, so we cannot say for certain whether these two contigs are a part of the same virus or if they represent two different Dicistroviruses.

We identified a contig corresponding to the *Secovirus* family in four bee species collected from 4 continents and Oceania. This virus shares its closest homology with viruses detected primarily in plant species<sup>62</sup>. While additional studies are needed to determine if this virus can replicate in bees, its broad distribution suggests that it is closely associated with bee species. Other recent studies have identified viruses with sequence similarity to plant viruses in bee samples<sup>26,27</sup>. In one case, for tobacco ringspot virus, replication was confirmed in bee samples<sup>27</sup>. It



remains to be determined if these viruses are truly plant-associated viruses that simply contaminate bees or if in fact bees are an important host. Regardless, the fact that bees carry plant-associated viruses so frequently suggests that they may be important vectors of plant diseases in the field. Indeed, recent studies found that *A. mellifera* are able to distribute plant viruses through pollination services<sup>63</sup>.

A study by de Miranda *et al.* characterized the first virus from the viral family *Tymoviridae* in *A. mellifera* honey bees, and also found a virus from this family replicating in *Varroa destructor* mites (Bee macula-like virus and *Varroa* Tymo-like virus, respectively)<sup>5</sup>. Our phylogenetic analysis suggested that the virus we observed is distinct from the previously described bee virus in this family, and is more closely related to viruses within the *Tymovirus* genus (as opposed to the *Maculavirus* genus for the other previously identified bee virus in this family). We found evidence of this Tymo-like virus in two distant geographic locations (New Zealand and United States), suggesting that this virus is widespread. Additionally, we found this virus in two different bee species (*Apis mellifera* and *Bombus impatiens*) indicating that this virus can infect both *Apis* and non-*Apis* bees.

Until recently, none of the viruses found to infect bees were (–)ssRNA viruses. However, Remnant *et al.* found four novel (–)ssRNA viruses, including a novel *Rhabdovirus* in *A. mellifera* honey bees from South Africa, Tonga, and the Netherlands<sup>6</sup>. This group proposed the name *Apis mellifera Rhabdovirus* (ARV). We also detected this same *Rhabdovirus* (>99% similarity) in *Apis mellifera* and *Bombus impatiens* samples from the US (Pennsylvania). Data on these sequences and detection of the virus in Israel honey bee populations has recently been published<sup>55</sup>. Given that the virus is found in two bee species, we recommend referring to it as Bee *Rhabdovirus* (BRV).

*Partitiviruses* have a dsRNA genome, and have been observed mainly in plants and fungi<sup>64</sup>. A recent study examining the invertebrate RNA virosphere<sup>56</sup> identified a number of *Partitiviruses* within animal hosts from a wide range of arthropod species, including insects. A phylogenetic analysis of our novel *Partitivirus* revealed that it is most closely related to this newly identified group of Arthropod *Partitiviruses*. The distribution of this virus, as detected here, suggests that it is also widespread, being found in Europe (Switzerland and Georgia) and the US (California). To the best of our knowledge, this is the first *Partitivirus* identified in bees, and further, this is the first evidence of a dsRNA virus in bees.

Viruses that utilize a DNA genome are very rare in bees, with very few being observed to this point<sup>26,65</sup>. Here, we identified a viral contig that is closely related to other members of the *Circoviridae* family of viruses, a viral family that utilizes a circular single stranded DNA (ssDNA) genome. These viruses have been found primarily within birds and mammals, but recent studies have begun to identify circular viruses within insects as well<sup>66</sup>, including 24 diverse circular viruses found to be infecting odonates (dragonflies and damselflies)<sup>67</sup>. It is possible that bees harbor a similarly diverse and robust set of circular viruses. The identified *Circovirus* was phylogenetically related to Dragonfly circularisvirus and to the best of our knowledge, represents the first evidence of a circular ssDNA virus observed in bees.

Overall, we found 127 unique contigs corresponding to novel viruses (i.e. viruses previously not observed in bees), with 27 contigs represented by >0.1% of the reads in a given sample (Tymo-like: 2 species & 2 regions, Seco-like: 4 species & 5 regions, Partiti-like: 1 species & 2 regions, Noda-like: 2 species & 2 regions, Dicistro-like: 2 species & 2 locations, Circo-like 1: 1 species & 2 regions, and Circo-like 2: 1 species & 1 region, see Supplementary Table 3), which thus represent important viruses for future analyses. Of these contigs, 7 contained RdRp or replicase regions that could be used for phylogenetic analyses (see Fig. 1). Our study suggests that, despite the focus on North American and European *A. mellifera* viral populations, these populations still remain to be fully characterized: indeed, the greatest number of newly identified viruses were found in North American samples. Additionally, when considering all the sample sets, it appears that all the newly identified viruses were found in multiple bee species (note that non-overlapping contigs for a novel dicistrovirus found in non-*A. mellifera* samples in India and *A. mellifera* samples in New Zealand – whether these represent distinct viruses or the same virus requires further sequence information). Finally, the viruses were largely evenly shared across the *A. mellifera* and non-*A. mellifera* bee species in a region, though overall, more sites (4 out of 7) had higher numbers of viruses identified in the *A. mellifera* samples than in the non-*A. mellifera* samples.

With our increasing awareness of the importance of viruses to bee health, and the complexity of bee viral ecology – where viruses are shared across insect and potentially plant species – there has been an extensive effort to fully characterize the bee viral community. Prior to 2010, more than 24 viruses were identified in bee populations<sup>2–7</sup>. Since then, several additional viruses have been identified using metagenomics approaches<sup>6,26,31</sup>. Together with our results, more than 30 viruses have been found in bees, however, this number likely represents a small subset of the viral diversity present in bee communities. Viral metagenomics for virus discovery in bees has thus far been fruitful, and our protocol has provided a foundation for future studies to continue identifying novel pathogens infecting global bee populations using an inexpensive, unbiased method for the sequence-independent detection of novel viruses. Additional studies are needed to confirm that these viruses replicate within bees, obtain the full viral genome sequences, examine any negative (or positive) impacts infection may have on bees, and examine how these viruses move within bee and plant communities. Our results suggest, however, that viral communities are generally shared globally across species, with species from geographically distant locations hosting similar suites of viruses and viral families.

## References

1. Brown, M. J. F. & Paxton, R. J. The conservation of bees: a global perspective. *Apidologie* **40**, 410–416 (2009).
2. Bailey, L. & Ball, B. V. *Honey Bee Pathology*. (Academic Press INC., 1991).
3. Runckel, C. *et al.* Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, Nosema, and Crithidia. *PLoS One* **6**, e20656 (2011).
4. Chen, Y. P. & Siede, R. Honey bee viruses. *Adv. Virus Res.* **70**, 33–80 (2007).
5. de Miranda, J. R. *et al.* Genome characterization, prevalence and distribution of a macula-like virus from *Apis mellifera* and *Varroa destructor*. *Viruses* **7**, 3586–3602 (2015).

6. Remnant, E. J. *et al.* A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations. *J. Virol.* JVI00158-17, <https://doi.org/10.1128/JVI.00158-17> (2017).
7. de Miranda, J. R. *et al.* Standard methods for virus research in *Apis mellifera*. *J. Apic. Res.* **52**, 1–56 (2013).
8. McMenamin, A. J., Brutscher, L. M., Glenn, W. & Flenniken, M. L. Abiotic and biotic factors affecting the replication and pathogenicity of bee viruses. *Curr. Opin. Insect Sci.* **16**, 14–21 (2016).
9. Klein, A.-M. *et al.* Importance of pollinators in changing landscapes for world crops. *Proc. Biol. Sci.* **274**, 303–13 (2007).
10. Ollerton, J., Winfree, R. & Tarrant, S. How many flowering plants are pollinated by animals? *Oikos* **120**, 321–326 (2011).
11. Tehel, A., Brown, M. J. F. & Paxton, R. J. Impact of managed honey bee viruses on wild bees. *Curr. Opin. Virol.* **19**, 16–22 (2016).
12. Singh, R. *et al.* RNA viruses in hymenopteran pollinators: Evidence of inter-taxa virus transmission via pollen and potential impact on non-*Apis* hymenopteran species. *PLoS One* **5**, (2010).
13. Levitt, A. L. *et al.* Cross-species transmission of honey bee viruses in associated arthropods. *Virus Res.* **176**, 232–240 (2013).
14. Fürst, M., McMahon, D. P., Osborne, J. L., Paxton, R. J. & Brown, M. J. F. Disease associations between honeybees and bumblebees as a threat to wild pollinators. *Nature* **506**, 364–6 (2014).
15. McMahon, D. P. *et al.* A sting in the spit: Widespread cross-infection of multiple RNA viruses across wild and managed bees. *J. Anim. Ecol.* **84**, 615–624 (2015).
16. Parmentier, L., Smagghe, G., de Graaf, D. C. & Meeus, I. Varroa destructor Macula-like virus, Lake Sinai virus and other new RNA viruses in wild bumblebee hosts (*Bombus pascuorum*, *Bombus lapidarius* and *Bombus pratorum*). *J. Invertebr. Pathol.* **134**, 6–11 (2015).
17. Potts, S. G. *et al.* Global pollinator declines: trends, impacts and drivers. *Trends Ecol. Evol.* **25**, 345–53 (2010).
18. Goulson, D., Nicholls, E., Botias, C. & Rotheray, E. L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science (80-.)*. **347**, 1255957–1255957 (2015).
19. Li, J. *et al.* The Prevalence of Parasites and Pathogens in Asian Honeybees *Apis cerana* in China. *PLoS One* **7**, 1–12 (2012).
20. Muli, E. *et al.* Evaluation of the distribution and impacts of parasites, pathogens, and pesticides on honey bee (*Apis mellifera*) populations in east Africa. *PLoS One* **9**, (2014).
21. Hou, C., Rivkin, H., Slabezki, Y. & Chejanovsky, N. Dynamics of the presence of Israeli acute paralysis virus in honey bee colonies with colony collapse disorder. *Viruses* **6**, 2012–27 (2014).
22. Malfroy, S. F., Roberts, J. M. K., Perrone, S., Maynard, G. & Chapman, N. A pest and disease survey of the isolated Norfolk Island honey bee (*Apis mellifera*) population. *J. Apic. Res.* **55**, 202–211 (2016).
23. Reynaldi, F. J., Sguazza, G. H., Albicoro, F. J., Pecoraro, M. R. & Galosi, C. M. First molecular detection of co-infection of honey bee viruses in asymptomatic *Bombus atratus* in South America. *Brazilian J. Biol.* **73**, 797–800 (2013).
24. Teixeira, E. W., Chen, Y., Message, D., Pettis, J. & Evans, J. D. Virus infections in Brazilian honey bees. *J. Invertebr. Pathol.* **99**, 117–119 (2008).
25. Granberg, F. *et al.* Metagenomic Detection of Viral Pathogens in Spanish Honeybees: Co-Infection by Aphid Lethal Paralysis, Israel Acute Paralysis and Lake Sinai Viruses. *PLoS One* **8**, (2013).
26. Schoonvaere, K. *et al.* Unbiased RNA shotgun metagenomics in social and solitary wild bees detects associations with eukaryote parasites and new viruses. *PLoS One* **11**, 1–18 (2016).
27. Li, J. L. *et al.* Systemic Spread and Propagation of a Plant-Pathogenic Virus in European Honeybees, *Apis mellifera*. *MBio* **5**, e00898–13 (2014).
28. Leathers, J. *Plebia Frontalis* (Friese): Stingless Bee. *California Department of Food and Agriculture*, <http://blogs.cdfa.ca.gov/Section3162/?tag=stingless> (2015).
29. Grozinger, C. M. & Robinson, G. E. The power and promise of applying genomics to honey bee health. *Curr. Opin. Insect Sci.* **10**, 124–132 (2015).
30. Mokili, J. L., Rohwer, F. & Dutilh, B. E. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2**, 63–77 (2012).
31. Levin, S., Sela, N. & Chejanovsky, N. Two novel viruses associated with the *Apis mellifera* pathogenic mite *Varroa destructor*. *Sci. Rep.* **6**, 37710 (2016).
32. Hunter, W. *et al.* Large-scale field application of RNAi technology reducing Israeli acute paralysis virus disease in honey bees (*Apis mellifera*, Hymenoptera: Apidae). *PLoS Pathog.* **6**, e1001160 (2010).
33. Liu, X., Zhang, Y., Yan, X. & Han, R. Prevention of Chinese sacbrood virus infection in *Apis cerana* using RNA interference. *Curr. Microbiol.* **61**, 422–8 (2010).
34. He, B. *et al.* Virome Profiling of Bats from Myanmar by Metagenomic Analysis of Tissue Samples Reveals More Novel Mammalian Viruses. *PLoS One* **8**, (2013).
35. Ng, T. F. F. *et al.* High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage. *J. Virol.* **86**, 12161–12175 (2012).
36. Sinha, R. *et al.* Index Switching Causes ‘Spreading-Of-Signal’ Among Multiplexed Samples In Illumina HiSeq 4000 DNA Sequencing. *bioRxiv*, <https://doi.org/10.1101/125724> (2017).
37. Andrews, S. FastQC: a quality control tool for high throughput sequence data, <http://www.bioinformatics.babraham.ac.uk/projects/> (2010).
38. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–20 (2014).
39. Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092 (2012).
40. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–52 (2011).
41. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
42. Sommer, D. D., Delcher, A. L., Salzberg, S. L. & Pop, M. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics* **8**, 64 (2007).
43. Bigot, D. *et al.* Discovery of *Culex pipiens* associated tunisia virus: a new ssRNA(+) virus representing a new insect associated virus family. *Virus Evol.* **4**, 1–13 (2018).
44. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410 (1990).
45. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, 501–504 (2007).
46. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
47. Adams, M. J. *et al.* Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Archives of Virology*, 10.1007/s00705-017-3358–5 (2017).
48. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–11 (2009).
49. Rombel, I. T., Sykes, K. F., Rayner, S. & Johnston, S. A. ORF-FINDER: A vector for high-throughput gene identification. *Gene* **282**, 33–41 (2002).
50. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, 36–42 (2013).

51. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
52. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
53. Rose, R., Constantinides, B., Tapinos, A., Robertson, D. L. & Prosperi, M. Challenges in the analysis of viral metagenomes. *Virus Evol.* **2**, vew022 (2016).
54. Sakamoto, J. M. *et al.* Bunyaviruses are common in male and female Ixodes scapularis ticks in central Pennsylvania. *PeerJ* **4**, (2016).
55. Levin, S. *et al.* Presence of Apis Rhabdovirus-1 in Populations of Pollinators and Their Parasites From Two Continents. *Front. Microbiol.* in press (2017).
56. Shi, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature* **540**, 539–543 (2016).
57. Pirk, C. W. W. *et al.* Statistical guidelines for *Apis mellifera* research. *J. Apic. Res.* **52**, 1–24 (2013).
58. Koch, H., Brown, M. J. & Stevenson, P. C. The role of disease in bee foraging ecology. *Curr. Opin. Insect Sci.* **21**, 60–67 (2017).
59. Mazzei, M. *et al.* Infectivity of DWV associated to flower pollen: Experimental evidence of a horizontal transmission route. *PLoS One* **9**, 1–16 (2014).
60. Traynor, K. S. *et al.* Multiyear survey targeting disease incidence in US honey bees. *Apidologie* **47**, 325–347 (2016).
61. Bonning, B. C. & Miller, W. A. Dicistroviruses. *Annu. Rev. Entomol.* **55**, 129–50 (2010).
62. Thompson, J. R., Kamath, N. & Perry, K. L. An evolutionary analysis of the Secoviridae family of viruses. *PLoS One* **9**, (2014).
63. Darzi, E. *et al.* The honeybee *Apis mellifera* contributes to Cucumber green mottle mosaic virus spread via pollination. *Plant Pathol.* 1–8, <https://doi.org/10.1111/ppa.12702> (2017).
64. Nibert, M. L. *et al.* Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. *Virus Res.* **188**, 128–141 (2014).
65. Gauthier, L. *et al.* The *Apis mellifera* filamentous virus genome. *Viruses* **7**, 3798–3815 (2015).
66. Rosario, K. *et al.* Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *J. Gen. Virol.* **93**, 2668–2681 (2012).
67. Dayaram, A. *et al.* Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. *Infect. Genet. Evol.* **30**, 278–287 (2015).

## Acknowledgements

We would like to thank Mario Padilla, Seymon Mukaku, Frederick Onyengo, Drew Wham and Briana Ezray for assistance with sample collection. This project was funded by a grant from USDA-APHIS (1515-8130-0501-CA) awarded to CMG and a National Geographic Young Explorers Grant (9893-16) to ZLF and JTK.

## Author Contributions

Project was conceived by D.A.G., Z.L.F. and C.M.G. Sample collection was facilitated by Z.L.F., A.B., M.F., M.W.G., K.M.K., J.T.K., S.D.K., O.L., E.M., S.S. and A.D.V. Molecular protocols for viral purification and nucleotide extraction were provided by J.M.S. Laboratory and molecular work was carried out by D.A.G. and Z.L.F., primer design and PCR validation performed by A.M.R., J.F.I.M., and C.R., and data analysis was done by D.A.G. and Z.L.F. Figures were generated by D.A.G., J.T.K. and H.M.P. The main text of the manuscript was written by D.A.G., Z.L.F. and C.M.G.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-27164-z>.

**Competing Interests:** The authors declare no competing interests.

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



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

© The Author(s) 2018