A Diverse Range of Novel RNA Viruses in Geographically Distinct Honey Bee Populations

Emily J. Remnant\textsuperscript{a,b,*}, Mang Shi\textsuperscript{c,d}, Gabriele Buchmann\textsuperscript{a}, Tjeerd Blacquière\textsuperscript{e}, Edward C. Holmes\textsuperscript{b,c,d}, Madeleine Beekman\textsuperscript{a,b} and Alyson Ashe\textsuperscript{d}

Behaviour and Genetics of Social Insects Laboratory, School of Life and Environmental Sciences, The University of Sydney, Australia\textsuperscript{a}, Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School, The University of Sydney, Australia\textsuperscript{b}, Charles Perkins Centre, The University of Sydney, Sydney, Australia\textsuperscript{c}, School of Life and Environmental Sciences, The University of Sydney, Australia\textsuperscript{d}, Wageningen Plant Research, Wageningen University & Research, Wageningen, The Netherlands\textsuperscript{e}.

Running Head: Novel RNA Viruses in Honey bees

*Address correspondence to Emily J. Remnant, emily.remnant@sydney.edu.au
Understanding the diversity and consequences of viruses present in honey bees is critical to maintain pollinator health and manage the spread of disease. The viral landscape of honey bees (Apis mellifera) has changed dramatically since the emergence of the parasitic mite Varroa destructor, which increased the spread of virulent variants of viruses such as Deformed wing virus. Previous genomic studies have focused on colonies suffering from Varroa and virulent viruses, which could mask other viral species present in honey bees, resulting in a distorted view of viral diversity. To capture the viral diversity within colonies that are exposed to mites, but do not suffer the ultimate consequences of the infestation, we examined populations of honey bees that have evolved naturally or been selected for resistance to Varroa. This revealed seven novel viruses isolated from honey bees sampled globally, including the first identification of negative-sense RNA viruses in honey bees. Notably, two Rhabdoviruses were present in three geographically diverse locations, and were also in Varroa mites parasitizing the bees. To characterize the antiviral response, we performed deep sequencing of small RNA populations in honey bees and mites. This provided evidence of a Dicer-mediated immune response in honey bees, while the viral small RNA profile in Varroa mites was novel and distinct from the response observed in bees. Overall, we show that viral diversity in honey bee colonies is greater than previously thought, which encourages additional studies of the bee virome on a global scale and which may ultimately improve disease management.

Honey bee populations have become increasingly susceptible to colony losses due to pathogenic viruses spread by parasitic Varroa mites. To date, 24 viruses have been described in honey bees with most belonging to the order Picornavirales. Collapsing Varroa-infected
colonies are often overwhelmed with high levels of picornaviruses. To examine the underlying viral diversity in honey bees, we employed viral meta-transcriptomics on three geographically diverse *Varroa*-resistant populations from Europe, Africa, and the Pacific. We describe seven novel viruses from a range of diverse viral families, including two viruses that are present in all three locations. In honey bees, small RNA sequences indicate that these viruses are processed by Dicer and the RNA interference pathway, whereas *Varroa* mites produce strikingly novel small RNA patterns. This work increases the number and diversity of known honey bee viruses, and will ultimately contribute to improved disease management in our most important agricultural pollinator.

**INTRODUCTION**

Viruses are some of the most common pathogens contributing to declining honey bee health and colony losses worldwide (1-4), and at least 24 viruses have been described in the western honey bee, *Apis mellifera* (1, 5-12). Historically, viruses were identified based on pathological symptoms present in diseased bees using serological methods (6, 7, 13, 14). Other than two DNA viruses (10, 11, 15), all currently characterized honey bee virus genomes comprise positive-sense, single-stranded RNA molecules (4, 16-19). Indeed, most viruses from honey bees fall into two families within the order *Picornavirales* (12, 16-21); the iflaviruses, including *Sacbrood virus* (SBV) and *Deformed wing virus* (DWV) (22, 23), and the dicistroviruses, including *Black queen cell virus* (BQCV) and *Acute bee paralysis virus* (ABPV) (24). More recently, genomic methods have identified additional viruses (5, 9, 10), which led to the discovery of a new genetic variant of DWV (23) and the *Lake Sinai virus* (LSV) group (5, 25). LSVs are common and widespread in honey bees, although whether they are associated with overt disease is unknown (26).
The prevalence, distribution and virulence of honey bee viruses seems closely associated with the ectoparasitic mite Varroa destructor (2, 3, 22-24, 27). Varroa can act as an important virus vector, causing a dramatic change in both viral landscape and virulence (3, 28, 29). As Varroa spread globally due to human-mediated translocation of honey bees during the early-mid 20th century, so did virulent viruses, leading to the widespread loss of managed and wild honey bee colonies (2, 3, 27, 29, 30). Viruses from the Picornavirales appear to have a particularly close association with Varroa (8).

While Picornavirales are commonplace, it is striking that negative-sense RNA viruses apparently seem to be absent from honey bee colonies. This is puzzling as negative-sense viruses, such as members of the Rhabdoviridae and Bunyaviridae, are widespread in other arthropods (31-33). Similarly, other categories of positive-sense RNA viruses, such as the Flaviviridae, are also common in insects but seemingly absent from those studies of honey bees performed to date (34, 35). The Flaviviridae are notable as they include mosquito and tick vector-borne viruses responsible for a number of important human diseases including dengue, West Nile and Zika as well as other insect-specific viruses (35, 36).

Honey bees exhibit multiple antiviral defense mechanisms, including one of the key innate immune responses in insects, the RNA interference (RNAi) pathway (19, 37, 38). During virus replication, double-stranded RNA intermediates are formed, which are recognized and cleaved by the endonuclease enzyme Dicer into 21-22 nucleotide fragments called small inhibitory RNAs (siRNAs; 39). These siRNAs are then bound by Argonaute proteins which guide the RNA-induced silencing complex (RISC) and degrade complementary RNA molecules such as the viral genomes (38). Honey bees produce siRNAs that match the...
predominant viruses in collapsing colonies (40), and can also produce a small RNA response when experimentally infected with double-stranded RNA (41, 42).

The use of RNA sequencing in a broad range of arthropods has greatly enriched our understanding of virus biodiversity (31, 33, 35, 43-46). It therefore seems opportune to revisit the viral diversity of honey bee colonies using similarly high powered techniques. Accordingly, we screened honey bee populations for the presence of RNA viruses using total RNA transcriptome sequencing, so called ‘meta-transcriptomics’ (33). We specifically focused on honey bee populations that have *Varroa* but do not appear to suffer any negative consequences (47), and populations without *Varroa*. These populations were selected to avoid the possibility that novel viruses are outcompeted by highly virulent viral strains associated with *Varroa*. Notably, we also sampled bees from geographically diverse locations in Europe, Africa and the Pacific to determine how viral diversity varies on a spatial scale.

**RESULTS**

**Viral diversity in Varroa-resistant and Varroa-free honey bee populations**

We examined the viral diversity of three *Apis mellifera* populations from Europe, Africa and the Pacific, by sequencing total ribosome-depleted RNA extracted from honey bee workers. We sampled honey bee colonies at the Amsterdam Water Dunes (The Netherlands), Stellenbosch and Robben Island (South Africa), and Vava’u and Tongatapu islands (The Kingdom of Tonga; Table 1). The colonies from the Netherlands are part of a selection program that started in 2008. These colonies are not treated against *Varroa*, but carry low numbers of mites (48). Colonies of *Apis mellifera capensis* in South Africa are similarly not treated against *Varroa* and are naturally resistant to mites (49). The honey bee population on Robben Island became infested with *Varroa* two years prior to sampling, and mite numbers
per colony remain low (49). In the pacific islands of Tonga, honey bees were most likely introduced during the 19th century and large numbers of feral colonies are found in multiple island groups. *Varroa* was introduced to the island of Vava’u in 2006. Due to the lack of commercial beekeeping, colonies are not treated to remove mites, and, like the South African population, honey bees appear to be naturally tolerant to *Varroa*. The honey bees on Tongatapu island have never been exposed to *Varroa* but are derived from the same original source population as those on the island of Vava’u (50). We synthesized libraries from pooled RNA extracted from five individuals per colony, and 100bp paired-end sequencing yielded a range of 4-9 Gb of data per library (Table 1). We assembled reads into contigs *de novo* using Trinity (51) and compared the resulting contigs to available viral protein sequences from GenBank with BLASTx.

We first examined the assembled contigs that matched to previously characterized honey bee pathogens (Table 1). We found similarities in the presence and absence of known viruses in all three *Varroa*-resistant populations. Contigs for Deformed wing virus (DWV), Black queen cell virus (BQCV) and Sacbrood virus (SBV) were present in all three locations. Acute bee paralysis virus (ABPV) was present in one colony from Robben Island, and Lake Sinai virus (LSV) was present in one colony each from the Netherlands and Tonga and two colonies in South Africa (Table 1). The LSV genomes from the Netherlands, Tonga and South Africa show 4 predicted open reading frames (ORFs) similar to previously characterized LSV-1 and -2 genomes (Figure 1). However, the nucleotide sequences from each location exhibit significant divergence, with 69-91% identity to previously characterised variants. We also examined our samples for contigs that matched to common fungal, bacterial and protozoan parasites of honey bees, such as the fungi *Nosema apis* and *N. ceranae*, the bacterial agents of European and American foulbrood, and the trypanosomes *Crithidia mellificae* and *Lotmaria*.
passim (4, 52). We observed contigs for *L. passim* in three colonies from Robben Island, South Africa, and one colony from the Netherlands (Table 1). The two colonies from mainland South Africa and one from the Netherlands contained contigs for *Nosema apis*, and one colony from Tonga contained contigs for *Nosema ceranae*, along with a single contig with similarity to *Leishmania sp* (Table 1).

Next we examined contigs that showed similarity to previously characterised virus sequences from positive-sense, negative-sense and double-stranded RNA genomes available in GenBank. We found genomic evidence of seven previously undescribed viruses in the honey bee transcriptomes, including four negative-sense and three positive-sense RNA viruses (Table 1-2, Figures 2-3). Among the negative-sense RNA viruses, two belong to the family *Rhabdoviridae* of the order *Mononegavirales* (Table 2, Figure 2A), and two belong to the family *Bunyaviridae* (Table 2, Figure 2B). Three positive-sense RNA viruses include one virus belonging to the family *Flaviviridae*, one related to Nora-viruses (*Picorna*-like) found in *Drosophila* (53), and one belonging to the family *Dicistroviridae* of the order *Picornavirales*, with homology to *Drosophila* C virus (Table 2, Figure 2 C-E). For the new viruses described here, individual abbreviations are derived from the host species (eg. *Apis mellifera*); followed by the name or category of virus and number if more than 1 (e.g. *Rhabdovirus*-1).

**Novel negative-sense RNA viruses**

*Rhabdo-like viruses*: *Apis mellifera* *Rhabdovirus 1* (ARV-1) has a 14,613 nucleotide (nt) genome with a prototypic *Rhabdovirus* structure (54) corresponding to the conserved gene order with five open reading frames (ORFs, Figure 2A). The most conserved ORF encodes for a 2,143 amino acid (aa) protein containing the RNA-dependent RNA polymerase (RdRp)
domain. The ARV-1 RdRp was most similar to Farmington virus (FARV), a virus originally
isolated from birds (55), with 30% aa identity (Table 2). The predicted G-protein of ARV-1
was also most similar to FARV (18%); however, BLAST searches of the other ORFs showed
no significant similarity to any known sequences. Apis mellifera Rhabdovirus 2 (ARV-2)
consists of a 14,029 nt genome, with five predicted ORFs (Figure 2A). The RdRp of ARV-2
was also related to FARV, but with much lower sequence similarity (23% aa identity, Table
2) compared to ARV-1. Unlike ARV-1, the ARV-2 G-protein showed structural and
sequence similarity to the hemagglutinin protein from Quaranfil virus from the
Orthomyxoviridae family (22% identity), which suggests this glycoprotein gene may be
acquired through an inter- (virus) family horizontal transfer event (33). Interestingly, both
ARV-1 and ARV-2 show evidence of being widespread. ARV-1 was found in six colonies
and in all three geographically diverse locations (Netherlands, South Africa and Tonga, Table
2 and 3), with high abundance ranging from 50 – 500 transcripts per million (Table 3). ARV-
2 was moderately abundant in South Africa and Tonga (5-17 TPM), and also detected in the
Netherlands (1.8 TPM, Table 2, Table 3). The ARV-1 and ARV-2 genomes from each
location exhibit 98-99% nt identity to each other (Figure 4). The RdRp protein sequences of
ARV-1 and ARV-2 formed a monophyletic group with FARV which were distantly related to
other members of the order Mononegavirales (Figure 3A, S1 Figure).

Bunya-like viruses: We identified two Bunya-like sequences, both from colonies from
Robben Island (Figure 2B). Bunyaviridae are negative-sense RNA viruses with three genome
segments (L- M- and S-), each containing a separate ORF (31). Due to lack of sequence
similarity to known sequences, for both viruses we could only identify the larger L-segments
encoding the RdRp domains. Apis mellifera Bunyavirus 1 (ABV-1) was present in three
colonies (Table 3). ABV-1 was most similar to a recently characterized class of
Bunyaviruses, the Leishbunyaviruses, identified in the insect trypanosomatid parasite *Leptomonas moramongo* (56). The RdRp protein of ABV-1 exhibits 56% aa identity to *LepmorLBV1* (Table 2). *Apis mellifera* Bunyavirus 2 (ABV-2) was present in one colony from Robben Island. The RdRp protein has 42% identity to *Wuhan Mosquito virus 1* (Table 2; 31). Phylogenetic analysis involving representative members from the *Bunyaviridae* placed ABV-1 in a clade with the *Leishbunyavirus* as a basal clade to other *Bunyaviruses* found within invertebrates (31, 56), and ABV-2 in the cluster of Phasmavirus-like *Bunyaviruses* (Figure 3B, S2 Figure).

Novel positive-sense RNA viruses

Flavi-like virus: *Apis mellifera* Flavivirus (AFV) was identified in one colony from Robben Island. The 20,414 nt positive-sense RNA genome contains a single ORF of 6,615 aa (Figure 2C). BLAST searches indicated AFV had 20% aa identity to *Gentian Kobu-sho-associated virus* (GKaV), a recently identified Flavi-like virus originally thought to be a dsRNA virus (Table 2; 35, 57, 58). Similar to GKaV and other newly identified Flaviviruses, the 20.4 kb AFV-1 genome length is longer than the typical length of previously characterized *Flaviviridae* (35). The phylogeny based on the RdRp/NS5 domain of AFV and other members of the *Flaviviridae* placed AFV in a clade of other recently discovered Flavi-like with large genomes (35; Figure 3C, S3 Figure), and Flavi-like virus segments identified in *Drosophila* species (Takaungu and Hermitage viruses; 34)

Dicistro-like virus: We identified a novel Dicistrovirus from one colony at the Amsterdam Water Dunes (*Apis mellifera* Dicistrovirus: ADV). The 9,126 nt genome contains two ORFs encoding the replication enzyme polypeptide and the capsid proteins, respectively, which is typical of dicistroviruses (Figure 2D). The polypeptide containing the RdRp exhibited the
highest genetic identity to *Drosophila C virus* (57% aa identity, Table 2). Phylogenetic analysis placed ADV to the same clade as *Cricket paralysis virus*, *Drosophila C virus*, and *Anopheles C virus* (Figure 3D).

**Nora-like virus:** Twelve separate contigs were assembled in one colony from Robben Island that each showed similarity to *Drosophila Nora virus* after BLASTx. To assemble a full-length genome, these contigs were ordered according to their closest related virus, *Drosophila pseudoobscura Nora virus*, and gaps were filled using RT-PCR and Sanger sequencing using primers spanning the neighbouring contigs. The resulting Apis mellifera *Nora virus* (ANV) partial genome sequence is 10,091 nt in length and covers the entire replicase, although it misses the first ORF at the 5’end of a typical *Nora virus* genome (Figure 2E; 53). In the phylogenetic tree, the ANV were closely related (52% ~ 54% aa identity) with *Nora viruses* isolated from different *Drosophila* species (Figure 3E).

**Small RNA profiles of ARV-1 and ARV-2 in honey bees**

One way of confirming that a putative virus genuinely infects the host from which it is sampled is the presence of an anti-viral immune response. In insects, likely candidates are the small RNA pathways that are utilized in viral defense (37, 40, 59). We therefore determined the presence of anti-viral small RNAs in bees infected with our novel viruses. We focused on ARV-1 and ARV-2 because they are the first negative-sense RNA viruses described in honey bees, and were present in colonies sampled from all three of our geographically diverse locations (Table 1).

We generated small RNA libraries from the abdomens of four *A. mellifera* samples: two from the Amsterdam Water Dunes and two from Robben Island. From each of these geographic
locations we used PCR to screen for individuals that were positive for ARV-1, and used one individual that tested positive for ARV-1 (AWD+ and RI+), and one that tested negative (AWD- and RI-). RI+ also tested positive for ARV-2. The four libraries were subjected to 50 bp single-end sequencing resulting in between 10 and 18 million reads per sample.

We mapped the resulting small RNA reads first to the *Apis mellifera* genome and then aligned the unmapped reads to the ARV-1 and ARV-2 genomes (Table 4). From this we found highly abundant small RNAs mapping to ARV-1 and ARV-2 (Table 5). Such small RNAs could either be random degradation products of viral RNA, or result from the honey bees’ antiviral immune response. Random degradation products of negative-sense RNA viruses would show a mixed size distribution of predominantly negative-sense fragments spanning the entire viral genome. In contrast, our small RNA reads have a size distribution of 21-22nt, occur equally in sense and antisense orientations, and map predominantly to the 5’ and 3’ ends of the ARV-1 and -2 genomes (Figure 5A-D). These features are typical signatures of Dicer-produced anti-viral RNAs, which occur when Dicer binds to a double stranded RNA intermediate and cleaves the double stranded RNA into viral siRNA (39).

The 5’ and 3’ genome bias suggests that replication intermediates at the ends of the ARV-1 and ARV-2 genomes provide dsRNA termini for Dicer to bind. Dicer-produced anti-viral RNAs occur at regularly spaced 21-23nt intervals (phases) starting at the dsRNA termini and fading with increasing distance from the termini, with a characteristic 2-3nt overhang (37). We looked for evidence of small RNAs occurring at regularly spaced intervals from the 5’ end of the ARV-1 and ARV-2 genomes using a phasing analysis (60). We detected a strong phasing signature for ARV-1 at 7 nt from the 5’ end, with a 2 nt overhang between sense and antisense strands (Figure 5E). We also detected a phasing signature for ARV-2, 2 nt from the
5’ end with an offset of 3nt between sense and antisense strands (Figure 5F). These data strongly indicate that Dicer is responsible for producing the 22nt small RNAs.

Finally, the antiviral immune response is also mediated by RNA-binding proteins such as Argonaute proteins, which bind to small RNAs and induce degradation of RNA sequences complementary to the small RNA (61). Argonautes often show a 5’ nucleotide preference in small RNA molecules (61), so we looked for nucleotide bias at the 5’ end of our small RNAs as compared to the base composition of the viral genome. Both the sense and antisense small RNAs against ARV-1 and ARV-2 display a highly significant reduction in 5’G and increase in 5’U as the 5’ nucleotide (Chi-squared test p<0.01 for both; Figure 5G, H).

Taken together, our data suggest that the ARV-1 and -2 small RNAs have been generated by Dicer acting on a double-stranded RNA replication intermediate, and that the small RNAs are subsequently bound by Argonaute proteins, indicating that the bees have an active anti-viral immune response against ARV-1 and ARV-2.

Small RNA profiles of ARV-1 and ARV-2 in mites

We next wanted to determine if ARV-1 and 2 are also found in mites feeding on infected bees. To that end we generated small RNA libraries from two V. destructor mites collected from A. mellifera individuals from the Amsterdam Water Dunes and Robben Island and performed Illumina 50bp single end sequencing, resulting in 18-20 million reads per sample (Table 4). We found small RNA reads mapping to ARV-1 and ARV-2 in both mite samples (Table 5, Figure 6).
The small RNA reads present in the mites show markedly different characteristics to the small RNA patterns in the honey bee samples. In mites, the antisense reads vastly outnumber the sense reads and span the length of the genome and have a distinct length distribution centered at 24 nt. Furthermore, the antisense reads correspond to the first four ORFs encoding the ARV-1 and ARV-2 viral proteins (Figure 6A-D). The sense reads, while much less abundant, show a broader size range than the antisense reads with peaks at 20 and 23 nt, and do not localize specifically to the ORFs (Figure 6E-F, S8 Figure). Phasing analysis from the 5’ end of the viral RNA did not show any evidence of phased RNAs of any size for in either sense or antisense reads (Figure 6G, H). Furthermore, there was only a weak 5’ nucleotide bias against 5’G and towards 5’U as compared to the base composition of the viral genome in sense reads for ARV-1 (Chi-squared test, P<0.05), no 5’ nucleotide bias for antisense reads for ARV-1 (Chi-squared test, NS), and no 5’ nucleotide bias for sense or antisense reads mapping to ARV-2 (Chi-squared test, NS) (Figure 6I, J). Lack of phasing, weak 5’ nucleotide bias and low quantity of sense reads suggest that the Dicer and Argonaute antiviral pathways do not act on replicating ARV-1 and ARV-2 dsRNA in the mites.

We also tested the 23-25nt long reads for signatures of the piRNA ping-pong amplification pathway, as this pathway has been implicated in viral defence in *Aedes aegypti* (62, 63). These signatures include a bias for uridine at the 5’ end and an adenine at position 10 on the complementary piRNA, as well as a ten nucleotide distance between the 5’ ends of overlapping sense and antisense reads (64). We generated a heat map showing the relative enrichment of each nucleotide at each position in the 23-25 nt long small RNAs. Although we could see some evidence of a U enrichment in position 1, there was no evidence to suggest enrichment of A at position 10. We also plotted the distance between the 5’ ends of overlapping 23-25nt long RNAs that map to ARV-1 and ARV-2 but found no evidence for a
peak at 10 nt. Thus, we could not detect any evidence for a ‘ping pong’ signature in the reads originating from either ARV-1 or ARV-2 in either mite sample (Figure 6 K, L and data not shown). Taken together, these data could mean that the small RNAs from ARV-1 and ARV-2 detected in mites are simply the products of random degradation, which is most likely the case for the sense small RNAs. However, in the antisense reads the strong bias for 24nt argues against random degradation.

The small RNA profiles in mites do not mimic those in honey bees, indicating that the mite reads do not come from ingested honey bee material. Thus, the high abundance of reads in the mites (Table 4), and the presence of sense reads (showing that a sense genome has been produced) suggest that both the mite and honey bee are *bona fide* hosts of ARV-1 and ARV-2.

**DISCUSSION**

We describe a diverse set of new viruses in *Varroa*-resistant or *Varroa*-free honey bee populations from three locations in Europe, Africa and the Pacific. We present genomic evidence of seven new RNA viruses, including three novel positive- and four novel negative-sense viruses. Our study therefore increases the number of known honey bee viruses from 24 to 31. *Apis mellifera* Rhabdovirus (ARV) -1 and -2 were found in three geographically distinct populations. Using small RNA sequencing we show that honey bees exhibit classic Dicer-mediated siRNA profiles suggesting an active bee immune response. We also report the first small RNA analysis in the mite *Varroa destructor* and show that ARV-1 and -2 are present in mites, although the small RNA profile is distinctly different from that in honey bees.
To our knowledge, this is the first identification of negative-sense viruses in honey bees. Three of the four novel negative-sense viruses (ARV-1, ARV-2, and Apis mellifera Bunyavirus (ABV-2) are related to viruses known to be present in insects (65). Our findings are thus consistent with recent studies describing the wide distribution of negative-sense viruses in arthropod hosts (31, 32, 66). Indeed, two negative sense viruses have also recently been found in the wild solitary bee Osmia cornuta, including one virus from the order Mononegavirales, and one from the family Bunyaviridae, indicating that viral diversity in other Hymenopteran species extends to negative sense viruses (67). One of the novel negative-sense viruses identified in our South African population from the Bunaviridae, ABV-1, was closely related to a recently isolated protist-infecting virus, the Leishbunyavirus LepmorLBV1 (56). LepmorLBV1 was isolated from the insect trypanosomatid parasite Leptomonas moramango, a parasite of Brachycera flies (68). Interestingly, the three colonies from Robben Island that contained ABV-1 contigs were also positive for the honey bee trypanosome, Lotmaria passim (52). We therefore cannot exclude that ABV-1 is a virus of protists that infect the bees.

The small RNA patterns of the Rhabdoviruses in honey bees show classical Dicer-mediated degradation profiles, providing strong evidence that ARV-1 and -2 viruses enter the cells of bees and begin to replicate. During viral replication, a double-stranded RNA replication intermediate is formed that can be recognised by the RNAi machinery and chopped by Dicer. The resulting small RNAs become part of an antiviral immune response (59). This strongly suggests that these novel negative-sense viruses are bona fide viruses capable of replication in honey bees.
In contrast, in *Varroa* the strong bias for 24 nt antisense small RNAs and lack of phasing suggests that replicating dsRNA is not a template for Dicer. Internal secondary structures within a negative-sense virus genome can provide dsRNA templates for Dicer (69, 70). An alternative Dicer-mediated anti-viral response seems the most parsimonious explanation for the small RNA profile found in the mites. Interestingly, however, the predominant size of 24 nt antisense reads is longer than expected for canonical Dicer products. It is unclear whether Dicer in *Varroa destructor* produces longer than usual RNA fragments, or if these 24nt RNAs are generated by a different viral degradation pathway. Other mites such as *Tetranychus* and *Metaseiulus* contain the components of the RNAi machinery, with considerable variation in gene copy number in Dicer and Argonaute proteins (71). An important step towards understanding the *Varroa* RNA interference pathway will be determining if the *Varroa* genome contains similar variation in key RNAi-mediating genes. It is intriguing that the 24nt antisense reads correspond to the ORFs of ARV-1 and -2. Similar “hotspots” have been observed before but their functional relevance is unclear (72).

Our study substantially expands the taxonomic diversity of honey bee viruses. Until now, most characterised honey bee viruses were restricted to the order *Picornavirales* (18, 19). The invertebrate-specific *Dicistroviridae* and *Iflaviridae* classes are evidently well adapted to parasitising insects, and many viruses in these groups show extremely broad host ranges (73-76), which may facilitate spread, allow viruses to replicate more readily in multiple hosts, and thus become more prevalent and easily detected. Positive-sense RNA viruses are also more abundant in eukaryotes generally (77), which likely contributes to the frequency with which they are detected in honey bees. Indeed, we found three novel positive-sense RNA viruses. Two of these viruses, *Apis mellifera* *Nora Virus* (ANV) and *Apis mellifera* *Dicistrovirus* (ADV), show relatively close evolutionary relationships with similar viruses from
Drosophila. Interestingly, early serological characterisation of Drosophila C virus (DCV) and the related Cricket-paralysis virus (CrPV) included honey bee samples for cross-reactivity to CrPV and DCV sera, and identified a honey bee variant (73, 78, 79). The third positive-sense virus identified here, Apis mellifera Flavivirus (AFV), follows from the recent identification of larger Flavivirus genomes (35), suggesting that arthropods could harbour a variety of viruses that will further illuminate the evolutionary origins of common viral categories (31, 35).

Until now viral surveys in honeybees predominantly focused on PCR based approaches which were heavily dependent on existing virus diversity (2, 3, 16, 25, 27), or on infectivity tests in honeybees, where viruses were identified based on their ability to multiply after injection into adult bees or pupae (18). This approach would necessarily exclude viruses that require different preparation methods or that are not amenable to crude extraction.

Importantly, our use of meta-genomic techniques were crucial in revealing a more complete bee virome, as many highly divergent viruses can only be detected at sequence level (43, 46). Our study was also aided by the recent explosion in novel virus genomes (31-33, 35, 45) that provide a more comprehensive database for BLAST searches. As this database continues to expand, it is likely that more new viruses in a range of host species will be isolated. Finally, our detection of novel viruses may in part reflect our focus on bee colonies that are resistant to, or free of Varroa, as the rapid spread of some viruses in the context of Varroa, most notably virulent variants of DWV, may have resulted in a general reduction in virus diversity.

Clearly, the discovery of the new viruses here suggests that the bee virome will continue to expand following more extensive metagenomic surveys in diverse geographic regions.

**MATERIALS AND METHODS**
Sample collection

Honey bee colonies and mites were sampled from Europe, Africa and the Pacific in 2013-2015. Seven *A. mellifera capensis* colonies were sampled from Robben Island, and five colonies from mainland South Africa in March 2013. *Varroa* mites were first identified in South Africa in 1997, and after initial deleterious effects, *A. m. capensis* colonies exhibited a natural mite resistance after 3-5 years (49). On Robben Island, *Varroa* mites were first detected two years prior to sampling (Allsopp, personal communication). In July 2014, ten colonies of *A. mellifera* were sampled from an apiary at the Amsterdam Water Dunes, The Netherlands, where *Varroa* has been present since the 1980s, and artificial selection for *Varroa*-resistant colonies has been ongoing since 2008 (48). In October 2015, feral and managed *A. mellifera* colonies were sampled from islands in the kingdom of Tonga. Nine colonies were sampled from Vava’u island, where *Varroa destructor* was introduced in 2006. Honey bees on Vava’u exhibit a natural tolerance that has enabled survival of colonies in the decade following the mite’s introduction. Four colonies were sampled from Tongatapu island where *Varroa* is not yet present. In Tonga, ten adult worker thoraxes per colony were crushed individually in 500µl of RNA-later (Qiagen) and transported at room temperature prior to storage at -80°C until processing. In Africa and the Netherlands, a minimum of ten adult worker honey bees were sampled per colony, frozen immediately on dry ice and stored at -80°C until processing.

Sample processing

Thorax and abdomen tissue of five adult bees from each colony (thorax only from Tonga) were homogenised in Trizol reagent (Thermofisher) and total RNA extracted following the manufacturer’s protocol. RNA from each individual was diluted to 200ng/µl, and 2.65ng RNA from each of the five individuals from each colony were pooled prior to DNAse
treatment (Ambion), followed by column purification (RNeasy mini kit, Qiagen). Total RNA was transported to the Australian Genome Research Facility (AGRF) on dry ice. Sample RNA integrity was confirmed using a Bioanalyzer (Agilent). RNA was subject to a ribosome depletion step (Ribo-Zero-Gold Human/Mouse/Rat) prior to preparation of Illumina TruSeq Stranded Total RNA paired-end libraries according to manufacturer’s instructions. Libraries were run on an Illumina HiSeq2000, 100bp paired-end sequencing for a total data yield of 4-9 Gb per sample.

Sequence read assembly and virus discovery

Sequencing reads were assembled de novo using Trinity (80). The resulting contigs were compared to reference protein sequences of all previously characterised viruses downloaded from GenBank using BLASTx (with an E value of 1-E5 to maximise sensitivity while minimising false positives; 35). Resulting virus-like contigs were then BLASTed to a non-redundant database to remove non-viral hits, such as host contigs with similarity to viral sequence. We also removed any contigs with high similarity to plant viruses which were more likely to be derived from food sources (although, there is a possibility that these could be replicating in bees, see: 81).

Virus sequences were aligned to the current databases of homologous viral proteins present in NCBI using MAFFT (82). Alignments were viewed and manually trimmed to remove large gaps and non-conserved regions, and further trimmed with TrimAL to remove ambiguously aligned regions (83). Maximum likelihood phylogenetic trees of each data set were inferred using PhyML (84), in parallel mode using Message Passing Interface (MPI) with 12 threads and 12 random starting trees. We used a best-fit model of amino acid substitution determined using ProtTest (85), a Subtree Pruning and Regrafting (SPR) branch-swapping
algorithm, and an approximate likelihood ratio test (aLRT) with the Shimodaira-Hasegawa-
like procedure to assess branch support.

Validation of novel viruses

We used PCR and sequencing to confirm the presence of each novel virus, designing primers
based on the contigs assembled from next generation sequencing. One virus, Apis mellifera
Nora-virus (ANV) produced fragmented contigs spanning an incomplete genome. Each
contig was ordered based on translated homology to the most related virus, Drosophila
pseudoobscura Nora virus, and primers spanning contig breaks were designed to confirm the
correct genome order and to sequence any unassembled regions (data not shown). We also
confirmed the arrangement of ARV-2 G- and L-proteins and the putative inter- (virus) family
horizontal transfer event, using PCR to amplify across the two ORFs and confirming with
Sanger sequencing.

We used Bowtie2 (86) to map reads to each of the novel virus genomes, Samtools to
determine the sequencing depth and coverage (87), and the RSEM program implemented in
Trinity to estimate the abundance of virus transcripts per million (TPM, Table 3; 51). Open
reading frames (ORFs, Figure 1) were annotated based on predicted amino acid sequences
that were longer than 200 nucleotides in length, as well as conserved position in the genome
compared to the most closely related viruses. Conserved domains were identified using NCBI
CDD BLAST searches (88).

Small RNA library preparation and sequencing

One μg of total RNA (prepared as above) from individual bees and Varroa mites was used to
generate a small RNA library using TrueSeq Small RNA kit (Illumina), following
manufacturers protocol. Samples were barcoded appropriately for pooling on an Illumina HiSeq2000, with 50 bp single end sequencing.

**Small RNA analysis**

Small RNA reads were quality checked, trimmed to remove the TruSeq adapter, and then mapped using CLC Genomics Workbench (Qiagen). They were first mapped to the *A. mellifera* genome allowing for up to two mismatches (length fraction = 1 and similarity fraction = 0.9) due to divergence between honey bee strains and the reference genome. The unmapped reads were subsequently mapped to ARV-1 and ARV-2 with the same stringency settings. The AWD and RI samples were mapped to the consensus viral genome generated by RNA sequencing (see above) from their geographic location. For further small RNA analysis (nucleotide size, genome position, base composition and 5’ read distance) the mapped read were exported as BAM files, indexed using Samtools (87) and then analysed in RStudio using viRome (89) and custom scripts.

**Phasing analysis of small RNAs**

To determine whether the small RNAs produced from the viral genomes showed evidence of Dicer phasing, we adapted an algorithm designed for detecting phased RNAs in plants (60). The logic behind this analysis is that if a RNA molecule is being cut every 22 nucleotides, there should be a relative accumulation of small RNA reads every 22 nucleotides. Thus, we summed the number of reads every x+22 nucleotides along the sequence (*P*) (where phase cycle position *(x)* = 1 to 22) using custom R scripts, then divided that by the number of remaining out-of-phase reads (*U*). Phasing scores were calculated for each phase cycle position in the 22nt window over 8 cycles as the signal is expected to degrade over time due to imprecision in the cut length. A relative accumulation of reads at a particular phase cycle
position will give a high phasing score. For the honey bee data we specifically looked for a phase cycle of 22, since that was the predominant size of the small RNAs. For the Varroa data we analysed four different phase cycle lengths (20, 22, 23 or 24 nucleotides).

\[
\text{Phasing score} = \ln \left( 1 + 10 \sum_{i=1}^{n} \frac{P_i}{1 + \sum U} \right)^{n-2}
\]

\(i\) = the phasing cycle; \(P_i\) = the number of small RNA reads at a given phase cycle position
\(U\) = the number of small RNA reads within the phase cycle, not at the phase cycle position (out of phase); \(n\) = the number of phase cycle positions occupied by at least one small RNA within the 8 cycles

Accession number(s)
All virus genome sequences generated in this study have been deposited in the GenBank database under accession numbers KY354230 to KY354234 and KY354236 to KY354244. Annotated protein sequence accession numbers are ARO50020 to ARO50067. Raw sequence data have been deposited to the Sequence Read Archive (SRA) (accession number SUB2176803; BioProject accession number PRJNA357165; BioSample accession numbers SAMN06140203 to SAMN06140219).

Acknowledgements
We thank Bruce White, Lamorna Osborne, Mike Allsopp and Benjamin Oldroyd for assistance with sample collection. We thank John Roberts for useful discussions of the results. This work was funded by the Australian Research Council (DP170100844 to MB and DE140100199 to AA), the Australian & Pacific Science foundation (APSF15-02 to EJR and MB) and the Marie Bashir Institute for Infectious Diseases and Biosecurity seed funding grant (to ER, ECH and MB). ECH is funded by an NHMRC Australia Fellowship.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors acknowledge the Sydney Informatics Hub and the University of Sydney’s high performance computing cluster Artemis for providing the high performance computing resources that have contributed to the research results reported within this paper.

REFERENCES

1. Ellis JD, Munn PA. 2005. The worldwide health status of honey bees. Bee World 86:88-101.

2. Mondet F, de Miranda JR, Kretzschmar A, Le Conte Y, Mercer AR. 2014. On the front line: Quantitative virus dynamics in honeybee (Apis mellifera L.) colonies along a new expansion front of the parasite Varroa destructor. PLoS Pathog 10:e1004323.

3. Martin SJ, Highfield AC, Brettell L, Villalobos EM, Budge GE, Powell M, Nikaido S, Schroeder D. 2012. Global honey bee viral landscape altered by a parasitic mite. Science 336:1304-1306.

4. Evans JD, Schwarz RS. 2011. Bees brought to their knees: microbes affecting honey bee health. Trends Microbiol 19:614-620.

5. Runckel C, Flenniken ML, Engel JC, Ruby JG, Ganem D, Andino R, DeRisi JL. 2011. Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, Nosema, and Crithidia. PLoS One 6:e20656.

6. Bailey L, Ball BV. 1991. Viruses, Honey Bee Pathology, Second Edition ed. Academic Press, London.

7. Allen M, Ball B. 1996. The incidence and world distribution of honey bee viruses. Bee World 77:141-162.
8. **Chen YP, Siede R.** 2007. Honey Bee Viruses, p 33-80, Advances in Virus Research, vol Volume 70. Academic Press.

9. **de Miranda J, Cornman R, Evans J, Semberg E, Haddad N, Neumann P,** **Gauthier L.** 2015. Genome Characterization, Prevalence and Distribution of a Macula-Like Virus from Apis mellifera and Varroa destructor. Viruses 7:2789.

10. **Gauthier L, Cornman RS, Hartmann U, Cousserans F, Evans JD, de Miranda JR, Neumann P.** 2015. The *Apis mellifera* filamentous virus genome. Viruses 7:3798-3815.

11. **de Miranda JR, Gauthier L, Ribiére M, Chen YP.** 2012. Chapter 8: Honey bee viruses and their effect on bee and colony health, p 71-102. In Sammataro D, Yoder JA (ed), Honey bee colony health: Challenges and sustainable solutions. CRC Press, Boca Raton, Florida.

12. **McMenamin AJ, Genersch E.** 2015. Honey bee colony losses and associated viruses. Current Opinion in Insect Science 8:121-129.

13. **Bailey L, Carpenter JM, Woods RD.** 1979. Egypt Bee Virus and Australian Isolates of Kashmir Bee Virus. Journal of General Virology 43:641-647.

14. **Bailey L, Woods RD.** 1977. Two More Small RNA Viruses from Honey Bees and Further Observations on Sacbrood and Acute Bee-Paralysis Viruses. Journal of General Virology 37:175-182.

15. **Bromenshenk JJ, Henderson CB, Wick CH, Stanford MF, Zulich AW, Jabbour RE, Deshpande SV, McCubbin PE, Seccomb RA, Welch PM, Williams T, Firth DR, Skowronski E, Lehmann MM, Bilimoria SL, Gress J, Wanner KW, Cramer RA, Jr.** 2010. Iridovirus and Microsporidian Linked to Honey Bee Colony Decline. PLoS ONE 5:e13181.
16. **Baker A, Schroeder D.** 2008. Occurrence and genetic analysis of picorna-like viruses infecting worker bees of *Apis mellifera* L. populations in Devon, South West England. *Journal of Invertebrate Pathology* **98**:239-242.

17. **Baker A, Schroeder DC.** 2008. The use of RNA-dependent RNA polymerase for the taxonomic assignment of Picorna-like viruses (order Picornavirales) infecting *Apis mellifera* L. populations. *Virology Journal* **5**.

18. **Ribière M, Ball B, Aubert MFA.** 2008. Natural history and geographical distribution of honey bee viruses. In Aubert M, Ball B, Fries I, Moritz R, Milani N, Bernardinelli I (ed), Virology and the honey bee. EEC Publications, Luxembourg.

19. **Brutscher LM, McMenamin AJ, Flenniken ML.** 2016. The Buzz about Honey Bee Viruses. *PLoS Pathog* **12**:e1005757.

20. Ed: Aubert M, Ball BV, Fries I, Moritz RFA, Milani N, Bernardinelli I. 2008. Virology and the honey bee. EEC Publications, Luxembourg.

21. **Le Gall O, Christian P, Fauquet CM, King AMQ, Knowles NJ, Nakashima N, Stanway G, Gorbalenya AE.** 2008. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-\( T = 3 \) virion architecture. *Archives of Virology* **153**:715-727.

22. **de Miranda JR, Genersch E.** 2010. Deformed wing virus. *Journal of Invertebrate Pathology* **103**:S48-S61.

23. **Mordecai GJ, Wilfert L, Martin SJ, Jones IM, Schroeder DC.** 2015. Diversity in a honey bee pathogen: first report of a third master variant of the Deformed Wing Virus quasispecies. *ISME J* **10.1038/ismej.2015.178**:1-10.

24. **de Miranda JR, Cordoni G, Budge GE.** 2010. The Acute bee paralysis virus-Kashmir bee virus-Israeli acute paralysis virus complex. *J Invertebr Pathol* **103**:S30-S47.
25. Cornman RS, Tarpy DR, Chen Y, Jeffreys L, Lopez D, Pettis JS, vanEngelsdorp D, Evans JD. 2012. Pathogen Webs in Collapsing Honey Bee Colonies. PLoS ONE 7:e43562.

26. Daughenbaugh KF, Martin M, Brutscher LM, Cavigli I, Garcia E, Lavin M, Flenniken ML. 2015. Honey bee infecting Lake Sinai viruses. Viruses 7:3285-3309.

27. Wilpert L, Long G, Leggett HC, Schmid-Hempel P, Butlin R, Martin SJM, Boots M. 2016. Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. Science 351:594-597.

28. Ryabov EV, Wood GR, Fannon JM, Moore JD, Bull JC, Chandler D, Mead A, Burroughs N, Evans DJ. 2014. A virulent strain of deformed wing virus (DWV) of honeybees (Apis mellifera) prevails after Varroa destructor-mediated, or in vitro, transmission. PLoS Pathog 10:e1004230.

29. Genersch E, Aubert M. 2010. Emerging and re-emerging viruses of the honey bee (Apis mellifera L.). Vet Res 41:54.

30. Rosenkranz P, Aumeier P, Ziegelmann B. 2010. Biology and control of Varroa destructor. J Invertebr Pathol 103, Supplement: S96-S119.

31. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, Chen L-J, Qin X-C, Xu J, Holmes EC, Zhang Y-Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. eLIFE 4:e05378.

32. Longdon B, Murray GGR, Palmer WJ, Day JP, Parker DJ, Welch JJ, Obbard DJ, Jiggins FM. 2015. The evolution, diversity, and host associations of rhabdoviruses. Virus Evolution 1.

33. 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 EC, Zhang Y-Z. 2016. Redefining the invertebrate RNA virosphere. Nature 540:539–543.
27

34. Webster CL, Longdon B, Lewis SH, Obbard DJ. 2016. Twenty-Five New Viruses Associated with the Drosophilidae (Diptera). Evolutionary Bioinformatics 12:13-25.

35. Shi M, Lin X-D, Vasilakis N, Tian J-H, Li C-X, Chen L-J, Eastwood G, Diao X-N, Chen M-H, Chen X, Qin X-C, Widen SG, Wood TG, Tesh RB, Xu J, Holmes EC, Zhang Y-Z. 2015. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the Flaviviridae and related viruses. J Virol 90:659-669.

36. Cook S, Moureau G, Kitchen A, Gould EA, de Lamballerie X, Holmes EC, Harbach RE. 2012. Molecular evolution of the insect-specific flaviviruses. The Journal of General Virology 93:223-234.

37. Ding S-W. 2010. RNA-based antiviral immunity. Nat Rev Immunol 10:632-644.

38. Brutscher LM, Flenniken ML. 2015. RNAi and Antiviral Defense in the Honey Bee. Journal of Immunology Research 2015:10.

39. Li Y, Lu J, Han Y, Fan X, Ding S-W. 2013. RNA Interference Functions as an Antiviral Immunity Mechanism in Mammals. Science 342:231.

40. Chejanovsky N, Ophir R, Schwager MS, Slabezki Y, Grossman S, Cox-Foster D. 2014. Characterization of viral siRNA populations in honey bee colony collapse disorder. Virology 454-455:176-183.

41. Maori E, Paldi N, Shafir S, Kalev H, Tsur E, Glick E, Sela I. 2009. IAPV, a bee-affecting virus associated with Colony Collapse Disorder can be silenced by dsRNA ingestion. Insect Molecular Biology 18:55-60.

42. Liu X, Zhang Y, Yan X, Han R. 2010. Prevention of Chinese Sacbrood Virus Infection in Apis cerana using RNA Interference. Current Microbiology 61:422-428.

43. Liu S, Vijayendran D, Bonning BC. 2011. Next generation sequencing technologies for insect virus discovery. Viruses 3:1849-1869.
44. Tokarz R, Williams SH, Sameroff S, Sanchez Leon M, Jain K, Lipkin WI. 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.

45. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui J-M, Bayne EH, Longdon B, Buck AH, Lazzaro BP, Akorli J, Haddrill PR, Obbard DJ. 2015. The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. PLoS Biol 13:e1002210.

46. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. Current Opinion in Virology 2:63-77.

47. Locke B. 2015. Natural Varroa mite-surviving Apis mellifera honeybee populations. Apidologie 47:467-482.

48. Blacquiere T. 2014. Natural selection for varroa resistant honeybee colonies, on Wageningen University & Research. https://www.wur.nl/en/newsarticle/Natural-selection-for-varroa-resistant-honeybee-colonies.htm. Accessed 2nd June.

49. Allsopp M. 2006. Analysis of Varroa destructor infestation of southern African honeybee populations. MSc. University of Pretoria, Pretoria.

50. Taylor B, Roper T. 2006. Honey bee disease survey and beekeeper training in Tonga. AgriQuality Limited.

51. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protocols 8:1494-1512.
52. **Schwarz RS, Bauchan GR, Murphy CA, Ravoet J, de Graaf DC, Evans JD.** 2015. Characterization of Two Species of Trypanosomatidae from the Honey Bee *Apis mellifera*: *Crithidia mellificae* Langridge and McGhee, and Lotmaria passim n. gen., n. sp. *Journal of Eukaryotic Microbiology* **62**:567-583.

53. **Habayeb MS, Ekengren SK, Hultmark D.** 2006. Nora virus, a persistent virus in *Drosophila*, defines a new picorna-like virus family. *Journal of General Virology* **87**:3045-3051.

54. **Walker PJ, Firth C, Widen SG, Blasdell KR, Guzman H, Wood TG, Paradkar PN, Holmes EC, Tesh RB, Vasilakis N.** 2015. Evolution of Genome Size and Complexity in the *Rhabdoviridae*. *PLoS Pathog* **11**:e1004664.

55. **Palacios G, Forrestor NL, Savij N, Travassos da Rosa APA, Guzman H, DeToy K, Popov VL, Walker PJ, Vasilakis N, Tesh RB.** 2013. Characterization of Farmington virus, a novel virus from birds that is distantly related to members of the family *Rhabdoviridae*. *Virol J* **10**:219.

56. **Akopyants NS, Lye L-F, Dobson DE, Lukeš J, Beverley SM.** 2016. A Novel Bunyavirus-Like Virus of Trypanosomatid Protist Parasites. *Genome Announcements* **4**:e00715-00716.

57. **Kobayashi K, Atsumi G, Iwadate Y, Tomita R, Chiba K-i, Akasaka S, Nishihara M, Takahashi H, Yamaoka N, Nishiguchi M, Sekine K-T.** 2013. Gentian Kobu-sho-associated virus: a tentative, novel double-stranded RNA virus that is relevant to gentian Kobu-sho syndrome. *Journal of General Plant Pathology* **79**:56-63.

58. **Bekal S, Domier LL, Gonfa B, McCopin NK, Lambert KN, Bhalerao K.** 2014. A novel flavivirus in the soybean cyst nematode. *J Gen Virol* **95**:1272-1280.

59. **Gammon DB, Mello CC.** 2015. RNA interference-mediated antiviral defense in insects. *Curr Opin Insect Sci* **8**:111-120.
60. Anonymous. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463:763-768.

61. Wilson RC, Doudna JA. 2013. Molecular Mechanisms of RNA Interference. Annual Review of Biophysics 42:217-239.

62. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, Monighetti C, Campbell CL. 2011. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiology 11:45.

63. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. 2012. Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. PLOS Pathogens 8:e1002470.

64. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. 2007. A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5' End Formation in Drosophila. Science 315:1587-1590.

65. Ed: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. 2012. Virus taxonomy: classification and nomenclature of viruses: Ninth report of the international committee on taxonomy of viruses. Press EA, Sand Diego.

66. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. 2015. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. Proceedings of the National Academy of Sciences 112:7536-7541.

67. Schoonvaere K, De Smet I, Smagghe G, Vierstraete A, Braeckman BP, de Graaf DC. 2016. Unbiased RNA Shotgun Metagenomics in Social and Solitary Wild Bees Detects Associations with Eukaryote Parasites and New Viruses. PLOS ONE 11:e0168456.

68. Yurchenko V, Votypka J, Tesarova M, Klepetkova H, Kraeva N, Jirku M, Lukes J. 2014. Ultrastructure and molecular phylogeny of four new species of monoxenous...
trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus
Wallaceina. Folia Parasitologica 61:97-112.

69. Ashe A, Bélicard T, Le Pen J, Sarkies P, Frézal L, Lehrbach NJ, Félix M-A,
Miska EA. 2013. A deletion polymorphism in the Caenorhabditis elegans RIG-I
homolog disables viral RNA dicing and antiviral immunity. eLife 2:e00994.

70. Sabin LR, Zheng Q, Thekkat P, Yang J, Hannon GJ, Gregory BD, Tudor M,
Cherry S. 2013. Dicer-2 Processes Diverse Viral RNA Species. PLOS ONE
8:e55458.

71. Palmer WJ, Jiggins FM. 2015. Comparative Genomics Reveals the Origins and
Diversity of Arthropod Immune Systems. Molecular Biology and Evolution 32:2111-
2129.

72. Jayachandran B, Hussain M, Asgari S. 2012. RNA Interference as a Cellular
Defense Mechanism against the DNA Virus Baculovirus. Journal of Virology
86:13729-13734.

73. Christian PD, Scotti PD. 1998. Picornalike viruses of insects, p 301-336. In Miller
LK, Ball LA (ed), The Insect Viruses, vol The Insect Viruses. Plenum Publishing
Corporation, New York.

74. Zhang X, He SY, Evans JD, Pettis JS, Yin GF, Chen YP. 2012. New evidence that
deformed wing virus and black queen cell virus are multi-host pathogens. Journal of
Invertebrate Pathology 109:156-159.

75. Singh R, Levitt AL, Rajotte EG, Holmes EC, Ostiguy N, vanEngelsdorp D,
Lipkin WI, dePamphillis CW, Toth AL, Cox-Foster DL. 2010. RNA viruses in
Hymenopteran pollinators: Evidence of inter-taxa virus transmission via pollen and
potential impact on non-Apis Hymenopteran species. PLoS One 5:e14357.
McMahon DP, Fürst MA, Caspar J, Theodorou P, Brown MJF, Paxton RJ. 2015. A sting in the spit: widespread cross-infection of multiple RNA viruses across wild and managed bees. Journal of Animal Ecology 84:615-624.

Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology 479–480:2-25.

Anderson DL, Gibbs AJ. 1988. Inapparent virus infections and their interactions in pupae of the honey bee (Apis mellifera Linnaeus) in Australia. J Gen Virol 69:1617-1625.

Plus N, Scotti PD. 1984. The biological properties of eight different isolates of cricket paralysis virus. Ann Virol (Inst Pasteur) 135 E:257-268.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644-652.

Li JL, Cornman RS, Evans JD, Pettis JS, Zhao Y, Murphy C, Peng WJ, Wu J, Hamilton M, Boncristiani HF, Zhou L, Hammond J, Chen YP. 2014. Systemic spread and propagation of a plant-pathogenic virus in European honeybees, Apis mellifera. MBio 5:e00898-00813.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30:772-780.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.
84. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. 
New algorithms and methods to estimate maximum-likelihood phylogenies: 
Assessing the performance of PhyML 3.0. Systematic Biology 59:307-321.
85. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of 
best-fit models of protein evolution. Bioinformatics 
doi:10.1093/bioinformatics/btr088.
86. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat 
Meth 9:357-359.
87. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, 
Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. 
Bioinformatics 25:2078-2079.
88. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, 
Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song 
JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. 
CDD: NCBI's conserved domain database. Nucleic Acids Research 43:D222-D226.
89. Watson M, Schnettler E, Kohl A. 2013. viRome: an R package for the visualization 
and analysis of viral small RNA sequence datasets. Bioinformatics 29:1903-1903.
FIGURE LEGENDS

Figure 1. Lake Sinai virus variants identified in this study. (A) Genome structure of LSV strains identified in the Netherlands (KY354242), Tonga (KY354241), and South Africa (KY354243 to KY354244), compared to previously characterised genomes of LSV-1 and LSV-2. Open reading frames are blue, and conserved functional domains are indicated. (Protein sequence accession numbers: ARO50053 to ARO50067.) (B) Maximum likelihood phylogenetic tree of nucleotide alignment of LSV strains from Netherlands, Tonga and South Africa, with LSV-1 and -2 and other strains described in (26).

Figure 2. Genome structure of novel viruses. (A) Rhabdo-like viruses, showing the genome size (nt) and Rhabdovirus open reading frames (N, P, M G and L protein/RdRp) of ARV-1 (KY354230 to KY354233) and ARV-2 (KY354234) relative to the previously characterized Farmington virus. (B) Bunya-like viruses, showing the identified L segment sizes (nt) and ORF of ABV-1 (KY354236) relative to LepmorLBV1, and ABV-2 (KY354237) relative to Wuhan Mosquito virus 1. (C) Flavi-like virus, AFV (KY354238) genome size (nt) and ORF relative to GKaV. (D) Dicistro-like virus, ADV (KY354239) showing genome size (nt) and two ORFs of ADV relative to Drosophila C virus. (E) Nora-like virus, ANV (KY354240) showing the putative 5’ truncated genome size (nt) and four ORFs relative to Drosophila Nora virus. (Protein sequence accession numbers ARO50020 to ARO50052).

Figure 3. Evolutionary relationships of novel viruses. Maximum likelihood phylogenies of (A) novel Rhabdoviruses ARV-1 and ARV-2; (B) novel Bunyaviruses ABV-1 and ABV-2; (C) novel Flavivirus AFV; (D) novel Dicistrovirus ADV; and (E) novel Nora-virus ANV. (Detailed trees of A-C can be found in S3-5 Figures).
Figure 4. *Apis mellifera* Rhabdovirus-1 variants. Maximum likelihood phylogenetic tree of nucleotide alignment of *Apis mellifera* Rhabdovirus 1 (ARV-1) variant genomes isolated from Netherlands (KY354230), South Africa (KY354231) and Tonga (KY354232).

Figure 5. Small RNA analysis of ARV-1 and ARV-2 in honeybees. (A) and (B) show the size distribution (15-37 nt) and 5’ nucleotide composition of small RNAs arising from ARV-1 (A) and ARV-2 (B). The sample from which the small RNA library was produced is shown above each graph. Bars plotted above the x-axis represent reads that map to the positive strand and those plotted below represent those that map to the negative strand. Bars are coloured according to the proportion of reads starting with A (red), C (green), G (blue), T (purple). (C) and (D) show the mapping of 20-23 nt long viRNAs to the genomes of ARV1 and ARV2 respectively. The cartoon shows the domains of the viral genomes as seen in Figure 2. (E) and (F) show the phasing score over 8 phasing cycles for each position within a 22nt phasing window. The upper and lower graphs show the phasing score for the sense and antisense reads respectively. A high score indicates many small RNAs fall into that phase position (indicated with arrowheads). This analysis was performed using the 21-23 nt long reads from (A) and (B). (G) and (H) show the 5’ nucleotide (Obs) compared with that expected (Exp) from the base composition of the viral genome for ARV-1 and ARV-2 respectively. Sense (S) and antisense (AS) were compared using a Chi-squared test. ** p-value <0.01.

Figure 6. Small RNA analysis of ARV-1 and ARV-2 in *Varroa*. Left hand side figures show ARV-1 and right hand side figures show ARV-2. A and B show the size distribution (15-37 nt) and 5’ nucleotide composition of small RNAs in mite 1 (M1) arising from ARV-1.
(A) and ARV-2 (B). Bars plotted above the x-axis represent reads that map to the positive strand and those plotted below represent those that map to the negative strand. Bars are coloured according to the proportion of reads starting with A (red), C (green), G (blue), T (purple). (C) and (D) show the mapping of 23-25 nt long viRNAs to the genomes of ARV1 and ARV2 respectively. The cartoon shows the domains of the viral genomes as seen in Figure 2. (E) and (F) show the size distribution and 5’ nucleotide composition of the sense small RNAs from (A) and (B), respectively, normalized to the number of sense reads present. They also show the mapping of the 19-24nt sense reads to the viral genomes. (G) and (H) show the results of phasing score over 8 phasing cycles for each position within a 24 nt phasing window for ARV-1 and ARV-2 respectively. The upper and lower graphs show the phasing score for the sense and antisense reads. This analysis was performed using the 24 nt long reads only. (I) and (J) show the 5’ nucleotide (Obs) compared with that expected (Exp) from the base composition of the viral genome for ARV-1 and ARV-2 respectively. Sense (S) and antisense (AS) were compared using a Chi-squared test. * p-value <0.05. (K) and (L) show the distance between the 5’ ends of overlapping reads on opposite strands (left panel of each) and the base composition of each nucleotide position (right panel of each) for the 23-25 nt long reads of ARV-1 and ARV-2, respectively.
**Table 1:** Data generated in this study and summary of the viruses identified.

| Location | Colony ID | Data generated<sup>a</sup> | Known Viruses | Novel Viruses<sup>d</sup> | Other |
|----------|-----------|-----------------------------|---------------|--------------------------|-------|
| The Netherlands | NE_AWD_1151 | 45393799 reads 9.08 Gb | DWV SBV BQCV LSV-NE<sup>c</sup> | ADV | Nosema apis |
| Amsterdam Water Dunes (AWD) | NE_AWD_1442 | 43418765 reads 8.69 Gb | DWV SBV BQCV | ARV-1 ARV-2 | Lotmaria passim |
| South Africa | SA_RI_A | 20107219 reads 4.02 Gb | DWV | ABV-1 ABV-2 ANV-1 | Lotmaria passim |
| Robben Island (RI) and Stellenbosch (SB) | SA_RI_11 | 20515230 reads 4.10 Gb | DWV | ABV-1 AFV-1 | Lotmaria passim |
| | SA_RI_49 | 18820078 reads 3.76 Gb | DWV SBV BQCV ABPV | ARV-1 ARV-2 ABV-1 | Lotmaria passim |
| | SA_SB_C1 | 44731233 reads 8.95 Gb | DWV SBV BQCV LSV-SA-1<sup>c</sup> | | Nosema apis |
| | SA_SB_K2 | 4202291 reads 8.41 Gb | DWV SBV BQCV LSV-SA-2<sup>c</sup> | | Nosema apis |
| Tonga Vava’u (V) and Tongatapu (T) | T_V9 | 18658353 reads 3.93 Gb | DWV SBV | ARV-1 | |
| | T_V10 | 21309419 reads 4.26 Gb | DWV | ARV-1 ARV-2 | |
| | T_T12<sup>a</sup> | 21141746 reads 4.23 Gb | DWV SBV BQCV LSV-TO<sup>c</sup> | ARV-1 ARV-2 | Nosema ceranae Leishmania sp. |
| | T_T23<sup>a</sup> | 19203423 reads 3.84 Gb | DWV SBV BQCV LSV-TO<sup>c</sup> | ARV-1 | |

<sup>a</sup> Varroa-free colonies; <sup>b</sup> 100bp paired-end Illumina HiSeq; <sup>c</sup> LSV variants presented in S1 Figure; <sup>d</sup> Novel viruses presented in Table 2 and Table 3.
**Table 2:** Classification, genome characteristics and geographic location of the novel viruses discovered in this study.

| Name             | Genome Order | Family         | Genome size (bp) | Closest relative (RdRp aa identity) | Geographic location | Isolated from |
|------------------|--------------|----------------|------------------|-------------------------------------|---------------------|---------------|
| ARV-1            | + ss RNA     | Mononegavirales Rhabdoviridae | 14,613           | Farmington Virus (30%)               | Netherlands         | Apis mellifera Varroa destructor |
| ARV-2            | + ss RNA     | Mononegavirales Rhabdoviridae | 14,028           | Farmington Virus (23%)               | South Africa       | Apis mellifera Varroa destructor |
| ABV-1            | + ss RNA     | unclassified Bunyaviridae | 6,032a           | Leishbunyavirus (56%)                | South Africa       | Apis mellifera |
| ABV-2            | + ss RNA     | unclassified Bunyaviridae | 6,496b           | Wuhan Mosquito Virus 1 (42%)         | South Africa       | Apis mellifera |
| AFV              | + ss RNA     | unclassified Flaviridae | 20,414           | Gentian Kobu-sho-associated virus (20%) | South Africa       | Apis mellifera |
| ADV              | + ss RNA     | Picornavirales Iflaviridae | 9,126            | Drosophila C virus (57%)             | Netherlands         | Apis mellifera |
| ANV-1            | + ss RNA     | unclassified Picorna-like | 10,091b          | Drosophila subobscura Nora virus (52%) | South Africa       | Apis mellifera |

a. Information from the reference.
b. Information from the reference.
### Table 3: Abundance and prevalence of novel viruses in sampled colonies.

| Novel Virus | Location       | Sample      | Abundance Estimation (TPM) | Average fold coverage |
|-------------|----------------|-------------|-----------------------------|-----------------------|
| ARV-1       | Netherlands    | NE_AWD_1442 | 47.75                       | 235                   |
|             | South Africa- RI | SA_RI_49     | 132.42                      | 571                   |
|             | Tonga- Vava’u  | T_V9        | 186.4                       | 705                   |
|             |                | T_V10       | 55.3                        | 263                   |
|             | Tonga- Tongatapu | T_T12      | 546.98                      | 3023                  |
|             |                | T_T23       | 348.21                      | 2232                  |
| ARV-2       | South Africa- RI | SA_RI_49     | 8.73                        | 62                    |
|             | Tonga- Vava’u  | T_V10       | 17.09                       | 86                    |
|             | Tonga- Tongatapu | T_T12      | 5.49                        | 32                    |
|             | Netherlands    | NE_AWD_1442* | 1.79                       | 5                     |
| ABV-1       | South Africa- RI | SA_RI_11     | 65.51                       | 398                   |
|             |                | SA_RI_49a   | 57.78                       | 337                   |
| ABV-2       | South Africa- RI | SA_RI_A      | 222.05                      | 1375                  |
| AFV         | South Africa- RI | SA_RI_A      | 30.9                        | 187                   |
| ADV         | Netherlands    | NE_AWD_1151 | 1.71                        | 13                    |
| ANV         | South Africa- RI | SA_RI_111'   | 1.53                        | 4                     |

* Multiple contigs formed (partial genome). TPM reported is the average for all contigs.
Table 4: Small RNA samples and data generated

| Sample name | Data generated (reads) | Number of reads mapped to genome | Percentage of reads mapped to genome | Number of unmapped reads mapped to genome | Number of unmapped reads mapped to novel viruses | Percentage of total reads mapped to novel viruses |
|-------------|------------------------|----------------------------------|--------------------------------------|-------------------------------------------|-----------------------------------------------|-------------------------------------------------|
| AWD+        | 10,395,269             | 3,101,145                        | 30%                                  | 7,294,124                                 | 80,473                                         | 1.1%                                             |
| AWD-        | 13,158,211             | 4,482,538                        | 34%                                  | 8,675,673                                 | 25                                             | 0.0%                                             |
| RI+         | 17,633,773             | 4,945,358                        | 28%                                  | 12,688,415                                | 1,565,107                                      | 12.3%                                            |
| RI-         | 12,454,373             | 885,663                          | 7%                                   | 11,568,710                                | 198                                            | 0.0%                                             |
| M1          | 18,673,943             | 4,077,137                        | 22%                                  | 14,596,806                                | 379,229                                        | 2.6%                                             |
| M2          | 20,061,865             | 4,246,819                        | 21%                                  | 15,815,046                                | 71,471                                         | 0.5%                                             |
Table 5: Number of small RNA reads mapped to ARV-1 and ARV-2

| Sample  | Apis mellifera Rhabdovirus 1 | Apis mellifera Rhabdovirus 2 |
|---------|-------------------------------|-------------------------------|
| AWD+    | 80,473                        | NA                            |
| AWD-    | 25                            | NA                            |
| RI+     | 1,550,604                     | 14,503                        |
| RI-     | 177                           | 21                            |
| M1      | 202,052                       | 177,177                       |
| M2      | 34,272                        | 37,199                        |
A. Lake Sinai virus ORFs

- LSV-1 (HQ871931)
- LSV-2 (HQ888865)
- LSV-Netherlands
- LSV-Tonga
- LSV-South Africa-1
- LSV-South Africa-2

B. Lake Sinai virus phylogenetic tree
A. Rhabdo-like viruses

Farmington virus (FARV)

Apis mellifera Rhabdovirus 1 (ARV-1)

Apis mellifera Rhabdovirus 2 (ARV-2)

B. Bunya-like viruses

Leptomonas moramango Leishbunyavirus (LepmorLBV1)

Apis mellifera Bunyavirus 1 (ABV-1)

Wuhan mosquito virus 1

Apis mellifera Bunyavirus 2 (ABV-2)

C. Flavi-like virus

Gentian Kobu-sho-associated virus (GKaV)

Apis mellifera Flavivirus (AFV)

D. Dicistro-like virus

Drosophila C virus

Apis mellifera Dicistrovirus (ADV)

E. Nora-like virus

Drosophila melanogaster Nora Virus

Apis mellifera Nora Virus (ANV)
