The discovery of Halictivirus resolves the Sinaivirus phylogeny

Diane Bigot,† Anne Dalmon,† Bronwen Roy, Chunsheng Hou,‡ Michèle Germain,† Manon Romary,† Shuai Deng,‡ Qingyun Diao,§ Lucy A. Weinert,‡ †James M. Cook,‡ Elisabeth A. Herniou† and Philippe Gayral†,‡

Abstract

By providing pollination services, bees are among the most important insects, both in ecological and economical terms. Combined next-generation and classical sequencing approaches were applied to discover and study new insect viruses potentially harmful to bees. A bioinformatics virus discovery pipeline was used on individual Illumina transcriptomes of 13 wild bees from three species from the genus Halictus and 30 ants from six species of the genera Messor and Aphaenogaster. This allowed the discovery and description of three sequences of a new virus termed Halictus scabiosae Adlikon virus (HsAV). Phylogenetic analyses of ORF1, RNA-dependent RNA-polymerase (RdRp) and capsid genes showed that HsAV is closely related to (+)ssRNA viruses of the unassigned Sinaivirus genus but distant enough to belong to a different new genus we called Halictivirus. In addition, our study of ant transcriptomes revealed the first four siniavirus sequences from ants (Messor barbarus, M. capitatus and M. concolor). Maximum likelihood phylogenetic analyses were performed on a 594 nt fragment of the ORF1/RdRp region from 84 siniavirus sequences, including 31 new Lake Sinai viruses (LSVs) from honey bees collected in five countries across the globe and the four ant viral sequences. The phylogeny revealed four main clades potentially representing different viral species infecting honey bees. Moreover, the ant viruses belonged to the LSV4 clade, suggesting a possible cross-species transmission between bees and ants. Lastly, wide honey bee screening showed that all four LSV clades have worldwide distributions with no obvious geographical segregation.

INTRODUCTION

The worldwide economic value of pollination is about €153 billion [1], as 70% of the main crops used for human consumption depend on insect pollinators [2]. Consequently, there is concern over the implications of recent declines in insect pollinators and raised awareness of the importance of honey bee (Apis mellifera) health. A combination of various elements, including pesticides, nutrition, management practices, environmental factors, parasites and pathogens, including viruses, have been linked to the decline of managed honey bees [3–11]. Honey bee colonies affected by colony collapse disorder (CCD) have been shown to host more pathogens than non-affected honey bee colonies [12]. However, whether pathogens are causing or contributing factors of CCD, or spread through opportunistic infections, remains unknown. Recently, in order to understand the causes of honey bee colony losses or deaths, extensive efforts have been made to monitor viruses of insect pollinators [13–15], reviewed by [16]. However, these efforts are hampered by limited knowledge of the true biodiversity of viruses infecting insect pollinators.

Since the discovery of the first honey bee viruses in the 1960s [17], 24 honey bee viruses and satellites have been described, reviewed in [18], and this number is increasing...
Several bee viruses are associated with CCD (reviewed in [14, 20]), but pathogenic effects per se are known for only a few of these viruses. One double-stranded DNA virus (Apis mellifera filamentous virus) has been described for honey bees [21, 22], but most bee viruses have positive single-stranded RNA [(+)ssRNA] genomes and belong to the Dicistroviridae and Iflaviridae families (Picornavirales). Other unclassified (+)ssRNA virus species have also been described, such as chronic bee paralysis virus (CBPV) and Lake Sinai virus (LSV), both showing similarities with members of the Nodaviridae family [23].

Between 2013 and 2015, 21–33 % of surveyed honey bee colonies were positive for LSV2 in the USA [24]. This high prevalence was further observed in 2013–2014 as over 34 % of colonies tested positive for pathogens in the Western US bore LSV infections [25]. Although LSV abundance is correlated with weak colonies, its pathology remains unknown and no visible symptoms have been attributed to LSV infection in honey bees [25]. Moreover, LSV has been detected in the Varroa destructor mites [25, 26] and a positive correlation with the presence of LSV and Nosema microsporidia has been demonstrated [24]. Furthermore, the few studies on LSV diversity and distribution were mainly based on American, Belgian and Spanish samples [12, 23, 25–32]. LSV has also been detected by PCR in Africa (Benin and Algeria) and South America (Colombia) but no sequences are available [9, 33, 34]. The limited geographic screening of LSV to date may well underestimate the true diversity of this virus. Currently, the International Committee on Taxonomy of Viruses (ICTV) recognizes only two LSV species, but other species or strains [25] have been described (Table S1, available in the online Supplementary Material).

Replicative forms of LSV, as demonstrated by detection of the negative-strand RNA intermediate by strand-specific PCR, have so far been found in only three bee species: in Apis mellifera, in which it was first discovered [23, 25, 26, 32], in the bumble bee Bombus pascuorum [30] and the solitary mason bee Osmia cornuta [26]. The presence of the replicative form of the virus in these species indicates that wild bees are probably natural hosts for LSV.

Here we report the discovery and description of a virus closely related to sinaiviruses in the sweat bee Halictus scabiosae, as well as the first detection of LSV in ants based on meta-transcriptomic analyses of wild Hymenoptera. We further collected honey bee flies from five countries and sequenced the ORF1/RNA-dependent RNA-polymerase (RdRp) region of LSV to study the genetic diversity and geographical distribution of the different sinaivirus clades.

RESULTS

Genome reconstruction of sinaivirus and new Halictivirus in wild Hymenoptera

A total of 580 million reads, 1.5 million assembled contigs and 1.2 million ORFs were analysed in this work (Table S2). Overall, seven new viral sequences (four complete genomes, one nearly complete and two partial) were found in six transcriptomes, all showing significant homology with (+)ssRNA LSV, a honey bee virus discovered in 2011 [23] (File S1).

The first three viral sequences labelled Halictus scabiosae Adlikon virus (HsAV) strains D, E and H were found in three individual sweat bees (sample IDs GA16D, GA16E and GA16H) sampled in Switzerland (Table S3). The HsAV genomes contain three ORFs: ORF1 of unknown function, ORF2 encoding the RNA-dependent RNA-polymerase (RdRp) and ORF3 encoding a capsid protein (Fig. 1). Full-length genomes were obtained for HsAV_D and E (detected in the individual transcriptomes of GA16D and GA16E). Their genomes were 5203 and 5238 nt in size respectively and were both highly covered by 34,837 reads (mean coverage 640.7 X) and by 14,108 reads (256.6 X), respectively (Fig. S1a and b). The 5201 nt genome of HsAV_H was nearly complete, covered by 533 reads (9.5 X) and filled with 207 undetermined nucleotides (Fig. S1c). All three HsAV sequences share over 96 % nucleotide identity (Fig. 2d). The genome of HsAV is smaller than that of the LSV2 species of Sinaivirus. It lacks ORF4 and harbours a repetition of 50 Adenine at the 5’ end of the genome (Fig. 1).

Four viral sequences were detected in ants. Two full-length genomes (LSV-Messor-R1 and LSV-Messor-R2) were found in a single Messor concolor harvester ant sampled in Crete, Greece (individual ID GA09R). Genomic organization was typical of LSV with four ORFs: ORF1 and ORF4 of unknown function, ORF2 encoding the RdRp and ORF3 the capsid (Fig. 1). The genomes were respectively 5816 and 5877 nt in size and covered by 21,642 (mean coverage 329.6 X) and 48,503 (741.8 X) reads for LSV-Messor-R1 and LSV-Messor-R2 respectively (Fig. S1d and e). Two additional partial sequences with homology to LSV were found in the ants M. barbarus (individual ID GA09J) and M. capitatus (individual ID GA09P). Both sequences were too small to be fully annotated (1613 and 553 nucleotides, respectively) but could be included in the ORF1/RdRp LSV phylogeny (see below).

HsAV has a specific genomic organization

The five new ant LSV and HsAV genomes were annotated and compared to the genomes of related viruses including LSV2 (Sinaivirus), anapheline-associated C virus (AACV; Chroparavirus), CBPV (Chroparavirus), and Mosinovirus (MoNV; Nodaviridae), Nodamura virus (NoV; Nodaviridae) (Fig. 1). The ant LSV-Messor sequences had the typical genomic organization of LSV2 (Fig. 1), in contrast with the HsAV sequences from sweat bees. HsAV has a type 3 RdRp domain (IPR002166) with a conserved catalytic domain (IPR007094) similar to those of Chroparavirus, Sinaivirus and some Nodaviridae (NoV) (Fig. 1). This suggests that HsAV has conserved the function of RNA virus replicase.
The ORF1 of LSVs contained a putative methyltransferase-guanylyltransferase (MTase-GTase) domain, also detected in Chroparavirus (AACV and CBPV), and in the N-terminal position of the protein A/RdRp of Nodaviridae (MoNV and NoV) [35, 36]. This MTase-GTase domain, with all conserved sites [35], could be identified by sequence homology in LSV-Messor-R1 and LSV-Messor-R2, but was lacking from all three HsAV sequences.

The capsid found in the HsAV genome was markedly different from those of LSVs (16% identity, 30% similarity (Blosum62) at the protein level). Sinaiviruses possessed a single short peptidase A21/N2 (IPR005313) domain at the 3' end of ORF3. A significantly longer peptidase A21/N2 domain was
found in the HsAV capsid ORF, but in an N terminal position. This peculiarity was also observed in the MoNV capsid. In addition, the HsAV and MoNV capsid ORFs displayed a second overlapping viral coat domain (IPR029053) in a C-terminal position. MoNV, which is a recombinant virus with a nodavirus-like RdRp, is the only other virus known to have such capsid domain organization. Other nodaviruses instead possess a viral coat domain (IPR029053) embedded with a peptidase A6 nodavirus coat domain (IPR000696) (Fig. 1).

Finally, the monopartite genomic organization of *Sinaivirus* and HsAV differs from the bipartite genomes of Chroparavirus and *Nodaviridae*, in which RNA1 encode ORF1 and RdRp, and RNA2 the capsid. This segmentation could explain some evolutionary dissimilarities observed between Chroparavirus/*Nodaviridae* and *Sinaivirus/HsAV*. Furthermore, segmentation could favour gene exchanges, possibly explaining the shared origin of the MoNV and HsAV capsids.

---

**Fig. 2.** Bayesian phylogenetic trees of ORF1, RdRp and capsid proteins of HsAV and new LSV sequences found in *Messor* ants. (a) ORF1 (603 amino-acid sites), (b) RdRp phylogeny (559 amino-acid sites). (c) Capsid phylogeny (884 amino-acid sites). HsAV sequences are indicated in orange and LSV sequences from ants in pink. GenBank accessions are indicated in the Table S6. Scale bar represents substitutions rate per site and node values are posterior probabilities. (d) Matrix of protein identities of RdRp (%) between shared sequences of full-length genomes.
Genome-scale phylogenies revealed the relationships of Sinaivirus, Halictivirus and Chroparavirus

Phylogenetic analyses were performed on each of the three ORFs (ORF1, RdRp and capsid) to explore the evolutionary history of the five new HsAV and LSV-Messor genomes.

First, phylogenetic analyses for ORF1 were performed from an alignment of 603 amino-acid sites using the LG+G+I evolutionary model (Fig. 2a). The phylogeny showed that the three HsAV strains formed a strongly supported monophyletic group (posterior probabilities=1), clearly distinct from the chroparavirus and sinaivirus clades. The two ant LSV sequences, LSV-Messor-R1 and LSV-Messor-R2 discovered together in a single ant, both belonged to Sinaivirus, which formed a monophyletic group (posterior probabilities=1).

Second, the RdRp phylogeny was built from a 559 amino-acid alignment using the LG+G+I evolutionary model (Fig. 2b). Since the RdRp gene is present in many RNA viruses, Nodaviridae sequences could be added as outgroups to root the LSV and HsAV clades. The RdRp phylogeny showed that HsAVs, which form a well-supported monophyletic group (posterior probabilities=1), and not the chroparaviruses, were the sister group of sinaiviruses (posterior probabilities=1). The RdRp tree also confirmed that ant LSV genomes belonged to the sinaivirus clade and did not form an independent lineage.

Third, the capsid phylogeny was built from an alignment of 884 amino-acid sites using the Blosum62+G+I evolutionary model (Fig. 2c). The monophyly of HsAV, the place of ant LSVs within the sinaivirus clade and the chroparavirus as outgroups were consistent with analyses of the other genes. Interestingly, the phylogeny showed that the capsid of MoNV derived from a common ancestor of the Halictivirus, and not from a sinaivirus. This evolutionary scenario was strongly supported by posterior probabilities of 0.99.

Analysis of RdRp protein identity showed that while the within-strain protein identity was high for HsAV (98.7%) and sinaivirus (83.2%), there was only 38% nucleotide identity between shared sequences from the sinaivirus and HsAV genomes (Fig. 2d), suggesting they could belong to different genera. This is a proposal that should be examined by the relevant ICTV committee to determine the appropriate species/genus demarcation criteria.

HsAV and LSV strain divergence

Nucleotide comparisons between full-length genomes of HsAV strains D, E and H revealed that the three sequences had accumulated 42 strain-specific single nucleotide polymorphisms (SNPs) widespread along the entire genomes. In HsAV_D, 14 synonymous and four non-synonymous SNPs were observed relative to the consensus sequence. The HsAV_E genome revealed two synonymous SNPs. HsAV_H contained four synonymous and 18 non-synonymous SNPs. This result suggests that the three assembled genomes are not contamination artifacts; for which nucleotide identity would be expected to approach 100%. The low level of polymorphism observed here confirmed that three different strains of a single virus species have been sequenced.

Molecular evolution suggests HsAV and ant LSVs are functional

Selective pressures acting on ORF1, RdRp and capsid of the newly discovered HsAV and ant LSV were estimated to verify if their evolutionary rates reflected those of functional infectious viruses.

As a comparison, we first estimated the non-synonymous over synonymous substitution (dN/dS) ratio typical of sinaivirus, chroparavirus and nodavirus to identify reference selective pressures acting on infectious viruses. On average, dN/dS ratios of the latter viruses were 0.17 for ORF1, 0.09 for RdRp and 0.01 for capsid gene. ORFs of ant LSVs were on average more constrained than in other infectious viruses: dN/dS=0.08 for ORF1 (likelihood ratio test (LRT): $\chi^2=2\Delta\text{LnL}=12.200$, $P=4.8E^{-4}$), $=0.03$ for RdRp (LRT, $\chi^2=29.173$, $P=6.6E^{-5}$), and $=0.01$ for capsid (LRT, $\chi^2=0.014$, $P=0.9$). RdRp of HsAV displayed similar selective constraints compared to other infectious viruses: dN/dS=0.03 for RdRp (LRT, $\chi^2=2.541$, $P=0.11$), whereas the ORF1 and capsid of HsAV seemed to evolve under a slightly more relaxed selection: dN/dS=0.3 (LRT, $\chi^2=1.082$, $P=0.3$) and=0.06 (LRT, $\chi^2=4.008$, $P=0.045$), respectively. Altogether, molecular evolution analyses of ant LSV and HsAV shows that dN/dS values were below 1 suggesting a selective regime of purifying evolution, as expected in functional infectious viruses.

Worldwide sinaivirus genetic diversity

There are currently 58 LSV sequences available in public databases, including six complete viral genomes, 41 sequences of the region overlapping ORF1/RdRp, three partial ORF1, two partial RdRp and six partial capsid sequences (Table S1). Most sequences were obtained from Apis mellifera, but a few come from the wild bees Andrena vaga, Bombus lapidarius and B. pascuorum. So far LSV sequences have been produced from only three countries: USA [12, 23, 25], Belgium [30–32, 37] and Spain [28, 29]. To increase both geographical and taxon sampling to improve the phylogenetic resolution, over 650 honey bees sampled worldwide were screened for the LSV ORF1/RdRp region. Thirty-six A. mellifera honey bees (pool or individual samples) were positive for LSV and were sequenced (Table S4). No LSV sequences from pooled honey bees produced electropherograms displaying double peaks that would indicate a mixture of different strains or species. LSV sequences were obtained from five new countries: eleven from France, three from Italy, five from Canada, five from China and seven from Australia, confirming that LSVs have a very wide geographic distribution across continents. In total 81 LSV sequences, as well as the three HsAV sequences were collated into a 594 nucleotide alignment of the ORF1/RdRp
A. mellifera in wasp Vespula pensylvanica virus logeny. The phylogeny distinguished at least four LSV clades and no evidence of geographical pattern could be associated with the four clades. The same pattern was observed at the continent scale as different LSV strains from three to four clades co-circulate in Europe, North America, Asia and Oceania (Fig. 3b). It should be noted that European LSVs from Belgium were overrepresented in clades A, C and D, reflecting a higher sampling effort in this country [26].

Interestingly, each country contained LSVs from multiple clades and no evidence of geographical pattern could be associated with the four clades. The same pattern was observed at the continent scale as different LSV strains from three to four clades co-circulate in Europe, North America, Asia and Oceania (Fig. 3b). It should be noted that European LSVs from Belgium were overrepresented in clades A, C and D, reflecting a higher sampling effort in this country [26].

The four LSV sequences associated with Messor ants all belong to clade B. The two virus genomes discovered within the same ants, LSV-Messor-R1 and R2, were phylogenetically distinct (posterior probabilities=0.99), while the two partial sequences from two other Messor ants, LSV-Messor-J and P, were closely related to LSV-Messor-R2 (posterior probabilities=1) (Fig. 3b). In ants, the two LSV clades formed two strains named LSV Messor 1 (comprising LSV-Messor-R1 sequence) and LSV Messor 2 (comprising LSV-Messor-R2, -J and -P sequences).

In this study, two bee samples (C004, C062) produced PCR amplicons using two different primer pairs (LSV and LSV-HsAV noted -LH in the phylogeny). For both C004 and C062 samples, overlapping sequences from both primer pairs were nearly identical (100 and 99.2 % nucleotide identity) and therefore clustered in the phylogeny, showing they result from a single virus population circulating in the bees.

**DISCUSSION**

Halictivirus: a new viral genus

The generation of metagenomic data via the development of next-generation sequencing (NGS) technologies has fuelled the discovery of many new viruses [38–40]. However, to date few honey bee viruses have been discovered this way. A novel flavivirus (+)ssRNA, the Moku virus, was recently found in A. mellifera, in the mite Varroa destructor and in the wasp Vespula pensylvanica [41]. In addition, NGS allowed the discovery of four new RNA viruses: Aphid lethal paralysis strain Brookings, Big Sioux River virus, LSV1 and LSV2 in honey bees [23]. In the era of metagenomics where genomic and phylogenetic analyses are powerful and efficient, new viral genomes deserve attribution of genus and species names, even in the absence of additional biological information, microscopic descriptions or pathology [42].

Our current study allowed the discovery and the description of three isolates of HsAV. The genomic reconstructions, annotation and phylogenies permit complete description of these new viruses, phylogenetically closely related to Choroparavirus, Sinaivirus and Nodaviridae. HsAV is distinguishable from other closely related viruses by the absence of the MTase-GTase domain within the ORF1, putatively implicated in 5’ cap formation [35, 36], and suggesting that it has another mechanism of initiation translation. The Adenine stretch at the 5’ end of the HsAV genome might form a non-conventional poly(A) head initiating virus translation, similar to the poly(A) head, which significantly enhances cap-independent mRNAs translation in some poxviruses [43]. Besides their specific genomic organization, several evolutionary features distinguish HsAVs. HsAVs form a monophyletic group, are genetically homogenous and clearly divergent from their closest relatives the sinaiviruses, from which they are separated by long branches of equivalent length to those defining the Choroparavirus and Sinaivirus genera. Altogether this supports the proposal that HsAVs belong to a distinct and new viral genus, which we call Halictivirus. Molecular evolution analyses revealed that all HsAV proteins are subjected to strong purifying selection, suggesting that this virus is functional and infectious. This is also suggested by a high transcriptome coverage (Fig. S1). However, symptoms associated to HsAV remain to be elucidated.

**LSVs infect multiple and diverse hosts**

LSVs were discovered in three independently collected harvester ants: LSV Messor 1 in Messor concolor and LSV Messor 2 in M. barbarus, M. capitatus and M. concolor. This is the first time this virus has been reported from insects outside the superfamily Apoidea. LSV was discovered in the honey bee Apis mellifera in North America [12, 23, 25], in Europe [26, 29, 32] and in Africa [33]. Moreover, LSVs have been detected in wild solitary bees of the Andrenidae family (Andrena vaga and A. ventralis), in Megachilidae (Osmia bicornis and O. cornuta) in Belgium [31], and in Apidae bumble bees in Colombia (Bombus atratus) [34] and in Belgium (B. lapidarius, B. pratorum and B. pascuorum) [30]. Ant LSVs formed a monophyletic group and were all unequivocally incorporated within LSV clade D. Given that all other known LSVs are from bee hosts, this result suggests that host jump events between bees and ants may have occurred. Interestingly, one ant harboured two viral strains, showing that co-infection might also occur in ants.

Since its discovery in 2011 [23], LSV screening in non-bee insects is lacking. However, other honey bee viruses have also been detected successfully in several other hymenopteran hosts, mostly non-Apis bees [16, 44]. Israeli acute paralysis virus (IAPV) has also been reported in the wasp Vespula vulgaris [45, 46] and the replicative form of the virus was found in Vespa velutina [47]. The invasive hornet V. velutina mainly feeds on honey bees [48], but detection of replicative viral genomes excludes a simple trophic contamination. Honey bee virus detection in ants is also scarce, and to our knowledge, CBPV and deformed wing virus (DVW) are the only honey bee viruses detected in...
Fig. 3. Bayesian phylogenetic tree of the ORF1/RdRp nucleotide region of all known and new LSV sequences. (a) Zoom of the LSV clade (594 nucleotide sites). New sequences from ants are indicated by the pink box. Samples from Europe (France, Italy, Belgium, Spain) are indicated by triangles, from North America (Canada, USA) by circles, from Asia (China) by squares and from Oceania (Australia) by stars. Taxons in bold were sequenced in this study. Red symbols are full-length LSV genomes. Taxon information for LSV sequences are in Table S1. (b)Collapsed phylogeny with HsAV as the outgroup. Scale bar represents the substitution rate per site. Node values are posterior probabilities.
ants. CBPV was found in *Formica rufa* (viral genome) and in *Camponotus vagus* (replicative genome) ants living close to apiaries [49]. The genome equivalent copy numbers of CBPV were comparable between ants and bees (10⁷ to 10¹¹ copy per individual) [49]. DWV was found in invasive Argentine ants *Linepithema humile* (replicative genome) [50]. DWV was found in New Zealand ants and a replicative form was found by strand-specific PCR in 7% of tested ants. Although no symptoms were observed in ants, the high copy number of the virus and the presence of viral replication suggest that honey bee viruses can infect ants. Additional studies, in which more samples should be analysed, are needed to determine if LSV infections in ants are dead-ends or could participate in spreading the viruses in bees or other insects.

The discovery of ant LSV would clearly benefit from further wider sampling and detection of the replicative form of the virus using specific detection of the minus-strand RNA genome [51]. However, several lines of evidence already suggest that *Messor* ants are not simple passive trophic carriers of LSV. First, *Messor* ants are mainly carnivorous, and dead bees would not be major foraging targets. Second, the *Messor* used for the transcriptome sequencing were not collected near apiaries. Third, ant LSVs were detected three times independently (i) in three ant species, (ii) sampled up to 2000 km apart and (iii) displaying high between-strains polymorphism exceeding Illumina sequencing errors; thereby excluding cross contaminations during the experiment. Altogether, these findings provide strong arguments in favour of a genuine LSV infection in ants.

Cross-species transmissions of viruses have been shown to occur more frequently than previously thought and play a major role in evolution compared to rare co-divergence events [52]. Adaptations of RNA viruses to a new host in a new environment are enhanced by high mutation rates and fast viral replication by RNA polymerases [53]. In addition, close phylogenetic relatedness between hosts may also facilitate cell entry via similar receptors [54]. Transmission vectors shared between host species can also mediate host switches. For instance, the DWV is transmitted by the *Varroa destructor* mite [51, 55–57] but also via the environment through contaminated pollen [58]. Both transmission routes mediate DWV inoculation in non-*Apis* hymenopteran species [46]. In the case of LSV, both pollen pellets and *Varroa* mite can carry LSV particles, but LSV replicative forms were absent from these vectors. In addition, LSV presence in the honey bee gut could indicate that a potential food-associated (i.e. through pollen) and/or fecal–oral horizontal transmission route can occur for LSV [25]. This kind of transmission appears more random in the case of the harvester ants, which principally eat seeds. Carnivorous ants (*Camponotus vagus*) were found to be potential hosts of CBPV, as replicating forms of the virus were found in ants living near infected apiaries [49].

**Resolution of sinaivirus phylogeny and characterization of LSV diversity**

By combining 47 separately published LSV sequences with 35 new LSV sequences from this work, the ORF1/RdRp phylogeny represents the most exhaustive characterization of LSV diversity so far. Furthermore, the use of the new Halicivirus as the outgroup allowed better resolution of the sinaivirus tree topology. No recombination was detected in this dataset, legitimating inferences drawn from this sinaivirus phylogeny. The phylogeny showed four main LSV clades, three of which correspond to the previously described clades A, C and D [26]. Clade B, which includes virus sequences from bees collected in the Northern hemisphere as well as in ants, is novel (Fig. 3). The ICTV currently recognizes two species within the new *Sinaivirus* genus: LSV1 and LSV2, respectively belonging to clades C and A. Our results suggest there are at least two additional LSV species corresponding to clades B and D, depending on the sequence divergence cut-off applied (Table S5). Previous work named some LSV sequences as LSV 5, but here multiple LSV 5 were dispersed in multiple clades and not corresponding to a distinct species. Altogether the phylogenetic analyses revealed the great diversity of sinaiviruses both in terms of species and strains, based on which taxonomic revision could be undertaken (Table S5).

Co-infections of a single host insect by distinct LSV strains or species was observed with the identification of both LSV Messor 1 and 2 strains (clade B) in a single *M. concolor* ant. Occurrence of LSV co-infections from clades A, C and D in single honey bees have also been recently reported from Belgium [26]. This shows that LSV co-infections are relatively frequent, whatever the level of relatedness between the viruses and whatever the hosts. As each species and each strain might have different pathology and virulence, this may complicate identification of symptoms associated with specific LSVs.

Strikingly, all four LSV clades have wide geographic distributions, revealed by our screening from several new countries. Moreover, all of the main clades were distributed across several continents. This confirms, on a far wider geographic scale, the observations based on LSV sequences from Belgium and the USA [26]. Recent honey bee trade such as import and export of queens or recurrent hive transports could explain the lack of geographical segregation of virus species. Notably, DWV also displays a global distribution of genotypes, reflecting a worldwide spread of viruses driven by *Varroa* mites [57, 59]. Interestingly, this heterogeneity in LSV and DWV genetic distribution contrasts with other bee viruses such as IAPV [60, 61], SBV [62] or black queen cell virus [63], for which genetic diversity shows clear biogeographic structure. As no symptoms have yet been associated with LSVs, which was only discovered in 2011, there is no regulation yet to manage LSV spread. Further research is required on the pathology of sinaiviruses and Halicivirus to determine their impacts on honey bee and wild pollinator health.
METHODS

Virus detection in bee and ant transcriptomes

The 43 transcriptomes used in this study were obtained from single adult insects (i.e. each individual was treated separately, without pooling) including 13 wild bees belonging to three Halictus species (Apoidea, Halictidae): H. saccarius, H. sexcinctus, H. simplex, and 30 ants from six species: Messor barbarus, M. concolor, M. structor, M. bouvieri, M. capitatus and Aphaenogaster subterranea (Formicidae). Twenty new transcriptomes were produced for this work to complement 23 previously published transcriptomes [64] (Table S3). Total RNA isolation of whole individual bees and ants was performed using standard protocols [65]. Succinctly, 50 nt single-end reads were produced by an Illumina Hiseq 2000 sequencer after cDNA synthesis using the SMART cDNA library Construction kit (Clontech, Mountain View, USA) from 5 µg of total RNA [64]. The 20 new transcriptomes were de novo assembled using the same method as previously [64] that is assembly with ABYSS V1.2.0 [66, 67] with Kmer set at 40 [68] and contig re-assembly with the CAP3 program [69].

ORFs were predicted on assembled contigs of the 43 transcriptomes using Prodigal V2_60 software for metagenomic data [70, 71] using the standard genetic code. Translated ORFs were annotated based on protein homology using the HHblits program implemented in the HHSuite package [72, 73]. To minimize false-positive results only ORFs displaying homology e-values <10^{-5} and probability >95 % were kept. Significant positive homology hits were then parsed to retrieve their NCBI taxonomic identifiers (TaxID; ftp://ftp.ncbi.nih.gov/pub/taxonomy) using the BLAST+program [74], and the corresponding taxonomic identification was assigned to the predicted ORFs. Viral ORFs were kept for further analyses. When multiple hits of the same viral family occurred in a single transcriptome, full-length viral genomes were reconstructed by assembly of the corresponding contigs into scaffolds (Geneious assembler program) and extension by successive mappings of initial reads (Geneious mapper program) using default parameters of Geneious 8.1.7 software [75]. A final mapping of all Illumina reads of initial transcriptomes was performed using the previously extended viral genome as a reference sequence to validate the accuracy of genome reconstruction and correct for mapping errors.

Genome annotation, phylogeny and molecular diversity of new viruses

In order to annotate new full-length viral genomes, conserved protein domains of all predicted genes were searched against the 14 protein domain databases available in the InterPro consortium [76] using InterProScan version 5 [77]. Multiple protein alignments were performed with MAFFT [78] using default parameters on ORF1, RdRp and capsid ORFs (Table S6). The best amino-acid substitution model was predicted using ProtTest [79]. Bayesian phylogenetic trees were inferred using MrBayes version 3.2.6 [80], by running four Markov chains for 10^6 generations. Branch support values indicate posterior probabilities estimated from trees sampled every 20 generations once the Markov chains had become stationary (determined by empirical checking of likelihood values).

By comparing complete aligned viral genomes, polymorphisms between HsAV strains were analysed. In order to assess the functionality of the three ORFs of HsAV and LSV associated with ants, dN/dS rates were estimated to quantify selection pressures. The PAL2NAL program [81] was first used to guide codon alignments using protein alignments. dN/dS were then estimated from codon alignments using branch-models [82, 83] of the CodeML program [84, 85] implemented in the PAML software version 4.9c [86]. Finally, different nested models were used to compare dN/dS of the branches of interest (newly discovered HsAV and ant LSVs) to those of reference viruses (LSVs, chroparavirus and Nodaviridae). Model comparisons were performed using LRTs, using χ^2 tests with type I error=0.05, degree of freedom=1 (i.e. the difference of the number of parameters between two models) and the test statistics χ^2 = 2ΔL.NL. (i.e. twice the difference of the log-likelihood of each model).

Large-scale de novo detection and phylogeny of LSV in honey bees

We screened for LSV sequences from 569 Apis mellifera honey bees sampled in France, Italy, Canada, China and Australia. Individual or pooled bees were sampled in the summers 2013–2016 (Table S4). Due to the worldwide sampling effort, biological material underwent distinct processes, summarized in Table S7 for simplicity. Briefly, individual or pooled bees were mechanically disrupted and homogenized in lysis buffer and total RNA was isolated according to kit manufacturer instructions using phenol/chlorophorm or guanidinium thiocyanate protocols. Total RNA was quantified using the Qubit Fluorometer or NanoDrop and 1–4 µg of total RNA was reverse-transcribed using random hexamer primers following the reverse transcription kit manuals.

PCR detection of multiple strains of LSV and/or sinaiviruses (LSV–HsAV) was performed using custom degenerate primers (Fig. 1, Table S8) targeting the region overlapping ORFI/RdRp, commonly used in LSV genetic studies [32]. PCR reaction mixes and cycling conditions (identical for LSV or LSV–HsAV primers) are detailed in Table S7. PCR products were analysed by electrophoresis in 1.5 % agarose gels, stained with GelRed and visualized under UV light. All positive PCRs were Sanger sequenced by GATC Biotech (Germany), Sangon Biotech (China) or the Hawkesbury Institute for the Environment (Australia) using forward M13FP and reverse M13-RP primers. Sequences from both strands were assembled using DNAMan software package, version 6.0.3 (Lynnon BioSoft, www.lynnnon.com). Electropherograms were manually corrected and ambiguities were replaced by N using Geneious R9 [75].
Multiple nucleotide alignments were performed with MAFFT [78] using default parameters on ORF1/RdRp sequences. In order to draw proper conclusions from the phylogeny, recombination was detected using the GARD program [87] implemented in the Datamonkey web server [88]. The best evolutionary model was predicted using ModelTest v2 [89]. A Bayesian phylogenetic tree was inferred using MrBayes version 3.2.6 [80] as described above.

Acknowledgements

We would like to thank Nicolas Galtier for data acquisition, Jonathan Romiguier for help in data accessibility, Marion Ballenghien for technical assistance and two anonymous reviewers for their comments and suggestions. Analyses largely benefited from the ISEM computing cluster platform. We are also grateful to the Genotoul Bioinformatics Platform Toulouse Midi-Pyrenees for providing computing and storage resources. We would like to thank the ARNIA company for Italian honey bee samples and ADAPI (Association pour le Développement de l’Apiculture) for honey bee samples from Southern France.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

35. Ahola T, Karlin DG. Sequence analysis reveals a conserved exten-
sion in the capping enzyme of the alphavirus supergroup, and a
36. Kuchibhatla DB, Sherman WA, Chung BY, Cook S, Schneider G
et al. Powerful sequence similarity search methods and in-depth
manual analyses can identify remote homologs in many appar-
37. Ravoet J, de Smet L, Wenseleers T, de Graaf DC. Vertical trans-
mision of honey bee viruses in a Belgian queen breeding pro-
39. Mokili JL, Rohwer F, Duttilh BE. Metagenomics and future perspec-
40. Rosario K, Breitbart M. Exploring the viral world through meta-
41. Mordecai GJ, Brettell LE, Parchori P, Villalobos EM, Martin SJ
et al. Moku virus; a new Iflavirus found in wasps, honey bees and
42. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR et al.
Consensus statement: virus taxonomy in the age of metagenom-
43. Shirokikh NE, Spirin AS. Poly(A) leader of eukaryotic mRNA
bypasses the dependence of translation on initiation factors.
44. Manley R, Boots M, Wilfert L. Emerging viral disease risk to polli-
nating insects: ecological, evolutionary and anthropogenic factors.
46. Singh R, Levitt AL, Rajotte EG, Holmes EC, Ostiguy N et al. RNA
viruses in hymenopteran pollinators: evidence of inter-taxon virus
transmission via pollen and potential impact on non-Apis hymen-
47. Yañez O, Zheng H-Q, Hu F-L, Neumann P, Dietemann V. A sci-
entific note on Israeli acute paralysis virus infection of Eastern hon-
eybee Apis cerana and vespine predator Vespula velutina. Apidologie
49. Celle O, Blanchard P, Olivier V, Schurr F, Cougoule N et al. Detec-
tion of chronic bee paralysis virus (CBPV) genome and its replica-
tive RNA form in various hosts and possible ways of spread. Virus
ants carry novel viruses in their new range and form reservoirs for
51. Yue C, Geners E, RT-PCR analysis of deformed wing virus in
honeybees (Apis mellifera) and mites (Varroa destructor). J Gen
52. Geoghegan JL, Duchène S, Holmes EC. Comparative analysis
estimates the relative frequencies of co-divergence and cross-
species transmission within viral families. PLoS Pathog 2017;13:
et al. Molecular approaches to the analysis of deformed wing virus rep-
lication and pathogenesis in the honey bee, Apis mellifera. Virol J
2009;6:221.
54. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS et al.
Cross-species virus transmission and the emergence of new epi-
Molecular approaches to the analysis of deformed wing virus rep-
ligation and pathogenesis in the honey bee, Apis mellifera. Virol J
2009;6:221.
56. Gisder S, Aumeier P, Geners E. Deformed wing virus: replica-
tion and viral load in mites (Varroa destructor). J Gen Virol 2009;
90:463–467.
Deformed wing virus is a recent global epidemic in honeybees
58. Mazzini M, Carrozza ML, Luisi E, Forzan M, Giusti M et al. Infectiv-
ity of DWV associated to flower pollen: experimental evidence of a
59. Berényi O, Bakonyi T, Derakhshifar I, Kögberger H, Topolska
et al. Phylogenetic analysis of deformed wing virus geno-
types from diverse geographic origins indicates recent global dis-
3611.
60. Chen YP, Pettis JS, Corona M, Chen WP, Li CJ et al. African acute
paralysis virus: epidemiology, pathogenesis and implications for
61.Palacios G, Hui J, Quan PL, Kalkstein A, Honkavuori KS et al.
Genetic analysis of Israeli acute paralysis virus: distinct clusters
6217.
62. Roberts JM, Anderson DL. A novel strain of sacbrood virus of
63. Reddy KE, Noh JH, Choe SE, Kweon CH, Yoo MS et al. Analysis
of the complete genome sequence and capsid region of black queen
cell viruses from infected honey bees (Apis mellifera) in Korea.
Population genomics of eusocial insects: the costs of a
vertebrate-like effective population size. J Evol Biol 2014;27:
593–603.
et al. Next-generation sequencing of transcriptomes: a guide to
RNA isolation in nonmodel animals. Mol Ecol Resour 2011;11:
650–661.
66. Birol I, Jackman SD, Nielsen CB, Qian JQ, Varhol R et al. De novo
transcriptome assembly with ABYSS. Bioinformatics 2009;25:
2872–2877.
ABYSS: a parallel assembler for short read sequence data.
68. Cahais V, Gayral P, Tsgakogeorga G, Melo-Ferrereira J, Ballenhgien
et al. Reference-free transcriptome assembly in non-model ani-
mals from next-generation sequencing data. Mol Ecol Resour
69. Huang X, Madan A. CAP3: a DNA sequence assembly program.
70. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW et al. Prodi-
gal: probkaryotic gene recognition and translation initiation site
71. Hyatt D, Locascio PF, Hauser LJ, Uberbacher EC. Gene and trans-
lation initiation site prediction in metagenomic sequences.
iterative protein sequence searching by HMM-HMM alignment. Nat
73. Söding J. Protein homology detection by HMM-HMM comparison.
BLAST+: architecture and applications. BMC Bioinformatics
2009;10:421.
Geneious basic: an integrated and extendable desktop software
platform for the organization and analysis of sequence data.


---

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.