Viruses of the invasive Argentine ants from the European Main supercolony: characterisation, interactions, and evolution

Lumi Viljakainen\textsuperscript{a}, Ida Holmberg\textsuperscript{a}, Silvia Abril\textsuperscript{b}, and Jaana Juvanuu\textsuperscript{a}#

\textsuperscript{a}Department of Ecology and Genetics, University of Oulu, Finland.
\textsuperscript{b}Department of Environmental Sciences, University of Girona, Spain.

#Corresponding author details: jaana.juvansuu@oulu.fi, +358 40 369 1441

\textbf{Keywords:} Argentine ant, \textit{Linepithema humile}, RNA viruses, invasive species, RNA-sequencing, virus evolution.

\textbf{Subject category:} 2.1. Insect RNA Viruses

\textbf{Word count:} 5798

\textbf{Depositories:} Genbank (accession number lists given on Table 1 and Table S1), Bioproject ID:PRJNA279338

\textbf{Abbreviations:} KBV, Kashmir bee virus; Lhu, Linepithema humile.
Abstract

The Argentine ant (Linepithema humile) is a highly invasive pest, yet very little is known about its viruses. We analysed individual RNA-sequencing data from 48 Argentine ant queens to identify and characterise their viruses. We discovered eight complete RNA virus genomes—all from different virus families—and one putative partial entomopoxvirus genome. Seven of the nine virus sequences were found from ant samples spanning seven years, suggesting that these viruses may cause long-term infections within the supercolony. Although all nine viruses successfully infect Argentine ants, they have very different characteristics, such as genome organisation, prevalence, loads, activation frequencies, and rates of evolution. The eight RNA viruses constituted altogether 23 different virus combinations, which, based on statistical analysis, were non-random, suggesting that virus compatibility is a factor in infections. We also searched for the virus sequences from New Zealand and Californian Argentine ant RNA-sequencing data and discovered that many of the viruses are found on different continents, yet some viruses are prevalent only in certain colonies. The viral loads described here most probably present a normal asymptomatic level of infection; nevertheless, detailed knowledge of the Argentine ant viruses may enable the design of viral biocontrol methods against this pest.
Introduction

The Argentine ant is one of the most successful invasive species, having spread from South America to all over the world during the past 160 years [1, 2]. It is among the International Union for Conservation of Nature’s 100 world’s worst invasive alien species, as the invasive populations displace native ant species, which has severe impacts on local biodiversity and ecosystem functions [3]. Argentine ants are unicolonial, meaning that the individuals between different nests mix freely. This lack of territoriality may lead to the formation of supercolonies when the number of nests is so large that direct cooperation between them is impossible [4]. In a supercolony, the nests belong to the same reproductive entity, genetic diversity is low, and aggression between individuals is absent. These characteristics are thought to explain the success of the invading populations [5], but they might also expose them to pathogens. The supercolonies are extensive; for example, one of the largest, called the “European Main supercolony”, spans 6000 km along the Mediterranean coast [6]. In contrast, the native supercolonies are comparatively small: from a few metres to several hundred [7, 8]. The life spans of introduced and native supercolonies can also differ greatly; the most successful introduced supercolonies are more than 100 years old, while native colonies last only a few years [8]. However, not all introduced colonies are long-lived: Argentine ant colonies in New Zealand have mean survival time of 14 years and it has been speculated that pathogens may have an effect on the population collapses [9, 10].

Viruses discovered from ant species are mostly single-stranded RNA viruses, though one single-stranded DNA virus has also been described [11–16]. Most in-depth virus research has been performed on the red imported fire ant (Solenopsis invicta Buren), to explore the possibility of using viruses as biopesticides against this invasive pest [17]. The positive-sense RNA virus Solenopsis invicta virus 1 (SINV1) was the first ant virus to be
characterized [18]. SINV1 has since been shown to preferentially infect red imported fire ants, at all host stages, reside mainly in alimentary canal and midgut, and to have the highest prevalence during the summer months [11, 12, 15, 18]. The red imported fire ant harbours six well-characterised viruses: five single-stranded RNA viruses (SINV1-5), and a single-stranded DNA virus: Solenopsis invicta denso virus (SiDNV) [11, 14, 20, 21]. SINV1 is associated with low initial weight in queens, SINV2 with low brood number, while SINV3 causes a decline in colony fitness in a laboratory environment [17, 22].

Argentinian, Australian, and New Zealand populations of the Argentine ant have been previously shown to replicate a dicistrovirus, Linepithema humile virus 1 (LHUV1), as well as three virus species identified first in honey bees: Deformed wing virus (DWV), Black queen cell virus (BQCV), and Kashmir bee virus (KBV) [23, 24]. Several other putative yet partial sequences of double-stranded DNA and single-stranded RNA viruses have also been found in New Zealand Argentine ant populations by taxonomic assignment of metatranscriptomic contigs [24].

In this study, we identified complete virus genomes from one of the largest Argentine ant colonies in the world: the European Main supercolony. We discovered genomes of known honey bee pathogen KBV, nine previously uncharacterised RNA viruses, and one entomopoxvirus. We classified the viruses by characterising their genome structure and by phylogenetic association with known viruses. We show that all of the viruses identified are producing a long-term infection in the European Main Argentine ant population, which enabled us to estimate the RNA viruses’ rates of evolution. We analysed the virus loads in 48 individual queens and show that each virus has specific expression and reactivation levels, and that the virus type combinations in them are not random. Finally, existing RNA-sequencing data from the European Main, New Zealand, and Californian supercolonies
allowed us to hypothesise which of the viruses are original Argentine ant viruses and which may have emerged during the invasion.

**Results**

**Virus identification**

Ants used in this study were collected in 2011 from the Argentine ant European Main supercolony in Catalonia and kept in the laboratory for three years until RNA sequencing in 2014. We assembled virus sequences from 48 RNA-sequencing samples, with either pooled or individual data, using Trinity assembler and Blastx searching against the National Center for Biotechnology Information (NCBI) virus protein database. The obtained contigs contained complete RNA virus genomes, but the DNA virus contig had only a coding region for entomopoxvirus rifampicin resistance protein. The virus sequences were deposited into the NCBI GenBank database (Table 1). We named the new RNA viruses according to their genomic organization (Fig. 1) and relatedness to known viruses by phylogenetic analysis (Fig. 2a-f): Linepithema humile polycipivirus 1 (LhuPcV1), Linepithema humile rhabdo-like virus 1 (LhuRLV1), Linepithema humile bunya-like virus 1 (LhuBLV1), Linepithema humile C-virus 1 (LhuCV1), Linepithema humile partiti-like virus 1 (LhuPLV1), Linepithema humile qinvirus-like virus 1 (LhuQLV1), Linepithema humile toti-like virus 1 (LhuTLV1), and Linepithema humile entomopoxvirus 1 (LhuEV1).

**The presence of the viruses in Argentine ant supercolonies and other insects**

Next, we studied whether the virus sequences found from the 2014 laboratory colony exist also in natural Argentine ant colonies by qPCR and searching existing transcriptome data sets. All the identified virus sequences were found from NCBI’s *L. humile* transcriptome shotgun assembly sequence data (TSA) produced from three queens and 15 workers (Bioproject ID = PRJDB4088) from the same European Main supercolony in Catalonia in
2012 [23] (Table 1). We also found two new complete virus sequences from the 2012 TSA data which were not in our laboratory colony: picorna-like virus (LhuPiLV1, LI759129.1) and narna-like virus (LhuNLV1, LI705470.1) (Fig. S1 and S2). Using virus-specific primers in PCR we found that ants (20 queens and 20 workers) collected in March 2018 from the same supercolony contained all the viruses except KBV and LhuCV1 (Table 1).

We made Trinity contigs using RNA-sequencing projects from New Zealand (PRJEB10079) and Californian (PRJNA46575) supercolonies to search for the new virus sequences. We found that KBV, LhuRLV1, LhuBLV1, LhuPLV1, and LhuTLV1 were also in a Wellington colony (Table 1). In addition, we identified a new polycipivirus (LhuPcV2, Fig. S1 and S2) from the New Zealand data. From the Californian data, we were able to identify only a putative LhuBLV1 capsid genome segment (EZ938627.1); however, we did identify a complete LHUV1 sequence (EZ933808.1) and a partial capsid of a new picorna-like virus (EZ934914.1). The names and Genbank IDs of all the complete virus genomes we discovered from the SRA data and a summary of their existence in European Main, New Zealand, and Californian supercolonies are in Table S1.

To understand if these viruses infect other insects, we searched the insect TSA database with the RNA virus genome sequences and found that only the Kashmir bee virus (KBV) is found in other insects: honey bees and several ant species (cut off range > 90% identity with > 90% query coverage by Blastp, Table 1). A virus similar to LhuQLV1 was found in the ant *Crematogaster osakensis* (IABS01051913.1, 71% identity with 62% query coverage).

*Kashmir bee virus*
KBV is a positive-sense RNA virus belonging to *Dicistroviridae* family (Fig. 1). The KBV genome sequence we found in the Argentine ants was 96% identical with 100% coverage to KBV isolated from honey bees (AY275710.1). KBV is already a well characterised virus species as it is believed to be involved in the global bee extinction [24].

**Linepithema humile polycipivirus 1**

LhuPcV1 is a positive-sense RNA virus which has been described previously as a member of a newly proposed polycistronic picorna-like RNA virus family: *Polycipiviridae* [27]. *Polycipiviridae* are mainly found in ant species, harlequin fly, and mixed insect samples identified by high-throughput RNA sequencing [27, 28]. The virus has four consecutive 5’ proximal ORFs and one long 3’ ORF (Fig. 1). The amino acid identity between the previously identified *Polycipiviridae* is low, and interestingly only ant polycipiviruses have an additional transmembrane protein part, encoded by a small ORF: ORF2b [27]. We found a new potential small ORF: ORF4b, which, according to the TMHMM transmembrane prediction programme, may have four transmembrane domains (Fig. 1). According to protein structure homology prediction with HHpred, ORFs 1, 3, and 4 are similar to picornavirus capsid proteins; ORF2, 2b, and 4b had no significant homology, while ORF5 is an RNA-dependent RNA polymerase (RdRp). We confirmed replication of LhuPcV1 in Argentine ant queens by quantifying the negative-sense RNA genome (Fig. S3).

**Linepithema humile C-virus 1**

LhuCV1 is a positive-sense RNA virus with a bipartite genome that contains six potential ORFs (Fig. 1). The 2487 nt genome segment contains three potential ORFs, where ORF1 and 3 are similar to hypothetical proteins of Anopheline-associated C virus (NCBI genome ID: 36225), with 40% identity and 95% coverage, and 28% identity and 73% coverage,
respectively. According to Blastp, ORF1 has an SP42 domain of putative membrane proteins of insect viruses. HHpred finds ORF3 protein sequence similar to a putative virion glycoprotein of insect virus (chroparavirus). Neither Blastp nor HHpred found sequences identical to ORF2.

The 3609 nt genome segment also contains three putative ORFs, of which ORF5 had no identical sequences by Blastp or HHpred. According to Blastp, ORF4 is 35% identical with 95% coverage to the hypothetical protein of Chronic bee paralysis virus (ID: 6188) and contains an RNA methyltransferase domain for capping viral RNA. ORF6 codes for RdRp with 47% identity and 99% coverage to Anopheline-associated C virus and according to HHpred the protein sequence is significantly similar to Dengue virus RdRp. According to the genome structure and phylogenetic analysis, LhuCV1 is most similar to Anopheline-associate C virus and Chronic bee paralysis virus, which form an unassigned virus family (Fig. 1 and 2). Tombus-like viruses which group together with LhuCV1 in the phylogenetic analysis, have very different genome structures and non-segmented genomes.

**Linepithema humile rhabdo-like virus 1**

LhuRLV1 is a negative-sense RNA virus of 11920 nucleotides. The LHuRLV1 genome contains six ORFs in all three reading frames (Fig 1). According to Blast, ORF1 is similar to the black garden ant (*Lasius niger*) glyceraldehyde-3-phosphate dehydrogenase (60% identity with 100% coverage); this sequence is probably wrongly annotated, instead indicating that the black garden ants have a similar virus. ORF2 is without any putative domains or similarity to known proteins according to a Blastp search. ORF3 codes for a small protein of 131 amino acids, which has no recognisable domains and is 27% identical with 91% coverage to Hubei rhabdo-like virus 3 (ID: 51928). ORF4 is 26% identical with 84% coverage to Hubei rhabdo-like virus 3 putative glycoprotein. ORF4 has a putative
transmembrane domain at the carboxy-terminal region, which may indicate that ORF4 encodes an envelope protein. ORF5 codes for a second small protein (133 amino acids) with no identifiable domains, yet with 35% identity (97% coverage) to ORFX of Berant virus (GenBank: KX580901.1) of unknown origin. ORF6 contains RdRp, mRNA capping domain V, and rhabdovirus L protein mRNA (guanine-N(7)-)-methyltransferase, and it is 68% identical to the Berant virus. According to HHpred prediction, ORF1 was similar to mononegavirales nucleoprotein (Borna disease virus, e-value 2.9 x 10^{-5}), ORF2 and 3 had no significant hits, and ORF4 had no significant hits but several low-level (e-values from 15 to 630) hits to viral fusion proteins. The small ORF5 had HHpred hits for possible alfa-helix coiled-coil domains of DNA binding proteins, such as c-jun and gal4.

The *L. humile* TSA database contained the LHuRLV1 genome in four different contigs, indicating that ORF1 produces one transcript, ORFs 2, 3, and 4 the second, ORF5 the third, and ORF6 the fourth transcript. All the contigs also contained a possible transcription termination signal sequence, ATTTAGAAAAA, at the 3′ end. Similar genome organisation to LHuRLV1 has been reported for Hubei Rhabdo-like virus 3 isolated from a Coleoptera beetle mix [26]. According to phylogenetic analysis, LHuRLV1 clusters among the other rhabdo-like viruses and is most closely related to the Hubei rhabdo-like virus 3 (Fig. 2).

**Linepithema humile bunya-like virus 1**

LHuBLV1 is a negative-sense RNA virus with an assumedly bipartite genome (Fig 1). Blastp identified the 6740 nt LhuBLV1 RdRp sequence as being 43% identical with 95% coverage to RdRp of Wuhan insect virus 16 (APG79216.1), while HHpred found the protein sequence similar to bunyavirus RdRp. We could not find the capsid segment from our host genome unmapped RNA-sequencing data but we did find a possible capsid segment (2170 nts) from the host genome mapped reads. The putative capsid sequence
was 25% identical with 83% coverage to Wuhan insect virus 16 and 100% identical with 100% coverage to L. humile uncharacterised protein (XP_012232778.1). The capsid sequence was also similar to other hypothetical proteins from several ant species (e.g. *Trachymyrmex zeteki*, *Vollenhovia emeryi*, *Trachymyrmex septentrionalis*, and *Ooceraea biroi*). The putative LhuBLV1 capsid segment was also found in the Californian TSA data (1162 nt, EZ938627.1). According to HHpred, the putative capsid protein sequence is similar to insect *phlebovirus* nucleocapsids (Granada virus: E-value 2.4e-48 and Buenaventura virus: E-value 2.4e-48) of the *Bunyaviridae* family.

According to phylogenetic analysis LhuBLV1 is most closely related to Wuhan insect virus 16 and Hubei bunya-like virus 3 (GenBank: KX884778.1), which both have bipartite genomes. Wuhan insect virus 16 and Hubei bunya-like virus 3 were isolated from arthropod and odonatan mix, respectively [26] (Fig. 2).

**Linepithema humile qinvirus-like virus 1**

LhuQLV1 is a bipartite negative-sense RNA virus with two putative ORFs (Fig. 1). *Qinviruses* are a new virus group proposed by Shi *et al.* in 2016 [28]. According to a Blastp search, the smaller genome segment of 1806 nt is 37% similar with 85% coverage to Wuhan insect virus 15 (ID:52673) and also similar to *Lasius niger* intron reverse-transcriptase maturase, the latter of which is probably a failed annotation, yet suggests that the black garden ants have a similar virus. The 5595 nt genome segment is 38% identical with 98% coverage to Wuhan insect virus 15 and HHpred finds significant protein similarity to vesicular stomatitis virus RdRp. According to the phylogenetic analysis, LhuQLV1 groups with other qinvirus-like viruses and associates most closely with Wuhan insect virus 15 [28] (Fig. 2).
**Linepithema humile partiti-like virus 1**

LhuPLV1 has a bipartite double-stranded RNA genome coding for the capsid and RdRp (Fig 1). The 1571 nt long capsid coding genome was 29% identical with 84% coverage to Hubei diptera virus 17 (ID: 52305) hypothetical protein (NC_033301.1). The RdRp coding segment of 1717 nt was 52% identical with 92% coverage to Hubei diptera virus 17 RdRp (NC_033302.1). HHpred revealed the RdRp protein sequence to be significantly similar to murine norovirus RdRp. Phylogenetic analysis clustered LhuPLV1 with Hubei diptera virus 17, which also has a bipartite genome [28] (Fig. 2). Other viruses clustering together with LhuPLV1 have varying numbers of genome segments [28].

**Linepithema humile toti-like virus 1**

LhuTLV1 is a double-stranded RNA virus of 7716 nt with two putative ORFs (Fig 1).

According to Blastp, ORF1 29% identical with 87% coverage to Shuangao toti-like virus (ID: 52063) hypothetical protein 1 (YP_009336731.1). ORF2 is 48% identical with 96% coverage to Leptopilina boulardi Toti-like virus (ID: 34144) RdRp (YP_009072448.1) and according to HHpred the protein sequence is similar to human picobirnavirus RdRp. In phylogenetic analysis, LhuTLV1 groups with other toti-like viruses and most closely to the Shuangao toti-like virus isolated from an insect mix [28] (Fig. 2).

**Linepithema humile entomopoxvirus 1**

Entomopoxviruses are large double-stranded DNA viruses of 200-300 kbp. We identified LhuEV1 from a contig coding for rifampicin resistance protein, then used it as a seed in NOVOplasty to further extend the virus genome using pooled *L. humile* DNA-sequencing data. We were able to construct a 47 kb contig coding for 48 ORFs with a minimum length of 100 amino acids. According to Blastp, 26 of the 48 ORFs were similar to known entomopoxvirus proteins. We decided to exclude the entomopoxvirus from further analysis.
due to it being very different from the other viruses and its low transcription levels, as almost five times fewer reads mapped to the rifampicin resistance protein contig than to the least abundant RNA virus (LhuTLV1). According to the reads mapped to LhuEV1 rifampicin contig, 39 of the 48 queens had detectable LhuEV1 rifampicin resistance mRNA.

**Linepithema humile RNA virus loads, prevalence, and interactions**

Viral RNA read amounts were analysed for each queen separately. Reads that did not map to the *L. humile* reference genome were mapped to viruses and we counted only reads that had a properly paired mate read also mapped to the virus sequence (Fig. 3a). The most common viruses were KBV, which was found in all 48 queens, and LhuPcV1, which was found in 41 queens (Fig. 3a, Table 2). The least common viruses were LhuQLV1, found in eight queens, and LhuTLV1 found in six queens (Fig. 3a, Table 2). The viral load variation in queens is presented by box plots for all the virus reads together (ALL) and for each virus separately (Fig. 3b). Although most of the virus read amounts varied over ten-fold, each virus clearly had a specific normal load level. For example, LhuRLV1 had the highest median level of about 4400 read units (around 20 000 actual properly mapped reads) and LhuTLV1 had the lowest median amount of two read units (around 10 actual properly mapped reads) per queen (Fig. 3a and b). The upper outlier values differ over 1.5 times of the interquartile value range from the third quartile value, thus indicating queens with unusually high virus loads. KBV and LhuPcV1 both had eight high outliers, LhuRLV1 had two high outliers and one low outlier (indicating unusually low virus load), and LhuPLV1 and LhuTLV1 had one high outlier each (Fig. 3b, Table 2).

On average, queens had three to four RNA virus types; at minimum, they had two virus types and at maximum six (Fig. 3c). All the queens infected with six viruses had the same
virus combination: five of the most abundant viruses KBV, LhuPcV1, LhuBLV1, LhuRLV1, and LhuPLV1, and the least abundant virus LhuTLV1 (Fig. 3c). The viruses formed 23 different combinations in the queens. To find out whether the combinations were random, we first compared the numbers of queens with the observed and expected virus combinations using the Fisher exact test. The Fisher exact test \( (p\text{-value}= 0.0052) \) suggested that the viruses were not randomly infecting the queens but that some virus combinations may be more or less common than expected. For example, the combination of KBV, LhuPcV1, LhuBLV1, LhuPVL1, and LhuCV1 was observed in five queens but expected only in one queen. However, the three most prevalent viruses, KBV, LhuPcV1, and LhuBLV1, were observed together in four queens and were expected to be found together in four queens, which indicates that the most prevalent viruses did not interfere with each other. Pairwise correlations indicated a negative correlation between LhuBLV1 and LhuQLV1 (Spearman’s correlation coefficient = -0.44, q-value= 0.042, Fig 3d). By random distribution, LhuBLV1 and LhuQLV1 would be expected to be found together in seven queens, yet they were never in the same ant. The other viruses did not have significant pairwise correlations in their distribution or read amounts (Fig. 3d).

**Linepithema humile RNA virus evolution**

To get an insight into the evolution of the Argentine ant RNA viruses, we took advantage of the existing *L. humile* virus contigs in the NCBI TSA database originating from samples collected in 2012. We calculated the total number of fixed mutations (substitutions) between the 2012 data and our 2014 virus sequences (from laboratory colonies originating from samples collected in 2011) (Table 2). The highest number of substitutions was in LhuPcV1, 424. The number of substitutions was highest in the positive-sense RNA viruses: KBV, LhuPcV1, and LhuCV1. Negative-sense and double-stranded RNA viruses had far fewer fixed mutations.
We then compared the substitution rates in the coding parts between the 2012 (TSA data) and 2014 (this study) virus genomes, by estimating the non-synonymous ($d_N$) and synonymous ($d_S$) substitution rates, i.e., substitutions that alter or do not alter the underlying amino acid, respectively (Table 2). LhuPcV1 had the highest $d_S$, 0.1685, which was an order of magnitude higher than for the other positive-sense RNA viruses, indicating a rapid rate of evolution. In contrast, the $d_N/d_S$ ratio of LhuPcV1 was the lowest, 0.02, which indicates that strong purifying selection has counteracted the deleterious effects of a high mutation rate on the underlying amino acid changes [29]. Interestingly, the $d_N$ values were almost identical (0.003) in all the three positive-sense RNA viruses, suggesting that this is a tolerable rate of nonsynonymous substitutions in highly mutating RNA viruses. For the other four viruses, the total number of substitutions was much lower and therefore the $d_N/d_S$ estimates are unreliable and not shown.

Discussion

The invasive Argentine ants are epidemiologically interesting as unlike most social animals, the physically separated nests in the massive supercolony are connected by the free movement of individuals, similarly to much of human civilization. This lack of territoriality may aid viral transmission within the colony.

From the European Main supercolony of the Argentine ants in Catalonia, we found eight highly expressing RNA virus genomes and one entomopoxvirus, all of which had established a long-term infection. Because we identified these viruses from both a long-term laboratory colony with specialized food (commercial cockroaches and honey) and wild colonies, we regard it likely that these virus sequences are from viruses specifically infecting Argentine ants and not from other ingested insects. Of these nine viruses, only
the honey bee pathogen KBV seems to infect other insect species [23, 30, 31], yet we did
find indications that viruses similar to LhuRLV1, LhuBLV1, and LhuQLV1 may infect other
ant species. KBV has previously been shown to infect the New Zealand supercolony of
Argentine ants, while the interaction between the ants and honey bees was shown to
increase their KBV loads [23, 24]. In our laboratory colony, KBV was the most prevalent of
the viruses, present in all the queens tested. According to the high viral load variability and
the number of queens with unusually high KBV loads (outliers), KBV seems to be very
active in these ants. In honey bees, stress caused by Varroa destructor mite infestations
has been shown to activate KBV replication [320]. High loads of SINV1 and several honey
bee viruses, among them KBV, have been shown to correlate positively with intra-colony
infection rate and pathology [14, 26, 33]. Therefore, the reactivity of KBV in the Argentine
ants might explain its high infection potential in ants, as well as in other insects. Although
KBV was present in Argentine ant samples collected from the European Main supercolony
in Catalonia in 2011 and 2012, we could not find it in the ants collected in March 2018. The
absence of KBV in March 2018 may reflect the lack of interaction between ants and bees
during winter, as well as KBV’s seasonal variation in honey bees, as the prevalence of
KBV in Europe is highest in spring and summer [34, 35]

We did not find in our samples dicistroviruses LHUV1 and BQCV, or iflavirous DWV, all of
which were identified from the New Zealand supercolony together with KBV [24].
Especially surprising was the absence of the LHUV1, since it has also been detected in
native Argentinian and introduced Australian colonies, and could hence be suspected to be
an indigenous Argentine ant virus [23]. Furthermore, we found the complete sequence of
LHUV1 in Californian Argentine ants but not in the 2012 TSA data of Catalonian Argentine
ants. The Australian, New Zealand, Californian, and European Main supercolonies most
likely originate from the same primary introduction event more than 100 years ago [1, 36,
The absence of LHUV1 from the European Main Catalonian supercolony may be due to a population bottleneck during the introduction, or alternatively, local European insect viruses may have displaced LHUV1. Argentine ants have been shown to carry many of the honey bee viruses and they might similarly be infected by other insect viruses. For example, LhuPcV1, LhuCV1, and LhuQLV1 were only found in the European main supercolony and not in New Zealand or California, suggesting that these viruses have either disappeared from the other two colonies or been transmitted after the European invasion (Fig. 4). However, validation of these speculations would require more extensive sampling of the European, New Zealand, and Californian Argentine ant colonies.

Although virus-virus interactions affect pathogenesis, little is known about them, especially between different virus species in nature [38–41]. A seasonal epidemic of one human respiratory tract-infecting virus species may shift and reduce the epidemics of the other virus species [42–44], with both negative and positive interactions between different respiratory tract viruses having been suggested [45]. Mathematical modelling of the respiratory tract viruses has suggested that they compete with each other mainly by infection speed, i.e., being the first to infect the cells, or by replicating quickly [46]. Ants provide an excellent model to study virus-virus interactions, because all the viruses infecting the whole organism can easily be isolated and sequenced. The eight high expressing RNA viruses in the Argentine ant queens made altogether 23 virus combinations from which the most common were combinations of i) KBV, LhuPcV1, and LhuPLV1 and ii) KBV, LhuPcV1, LhuBLV1, LhuPLV1, and LhuCV1. Both combinations were found from five queens (Fig. 4). Although our results suggested that the virus associations might not be random, especially for the least prevalent viruses, we could not see any significant positive correlation among the viruses by pairwise comparison. However, we did find significant negative correlation between LhuBLV1 and LhuQLV1, as
they were never found from the same queen. LhuBLV1 and LhuQLV1 are both bipartite negative-sense RNA viruses and hence they could compete for the same cellular process for replication if infecting the same cells.

The viruses with positive-sense RNA genomes had accumulated most nucleotide substitutions in comparison to the other RNA viruses. In particular, LhuPcV1 had over three times as many substitutions per nucleotide as the other two positive-sense RNA viruses. Apart from just genome size, substitution rates are affected by many viral life style factors, such as recombination rates, chronic versus acute infection (seasonality of epidemics), replication rate, pressure from host’s antiviral defence mechanisms, cell tropisms, and the existence of a secondary host [47–49]. KBV and LhuPcV1 had the highest rates of evolution as measured by the rate of synonymous substitutions (dS), indicating very different infection dynamics for these two viruses compared to the others. Both KBV and LhuPcV1 infected almost all of the queens and they both had high percentage of hosts with unusually high viral loads, indicating productive infection. LhuPcV1 had especially high outlier read values, over 1000 times higher than its median read value, which may indicate that once the virus is activated its replication speed is faster compared to the other viruses. It has been suggested that the fidelity of RNA virus replication may be adjusted for replication speed, to enable an optimal life style for the virus [50, 51]. Virus quasispecies theory suggest that viruses that have fast but error-prone replication produce a population of variants (a mutant cloud) that is able to adapt to a multitude of environmental changes, such as a completely new host species, while retaining a so-called master sequence, that has the highest fitness in the current environment [52]. Consistently with this theory, although LhuPcV1 produces many variants, it has very low ratio of non-synonymous to synonymous substitutions, indicating a
strong purifying selection to maintain the coding potential of the original genome, which is perhaps best suited for infection of the Argentine ants.

Our method misses viruses without poly-A-tails and viruses with low or restricted infection levels, hence there may be many more viruses still to be discovered, even from the Argentine ants of the European Main supercolony in Catalonia. In future, it will be interesting to discover how widely these viruses have spread throughout the massive supercolony, how the surrounding native insect viruses have affected the ant virome, and how the virus ecology influences the invasive prowess of the Argentine ant.

**Methods**

**Ants**

The Argentine ant samples were collected in April 2011 from the European Main supercolony in Catalonia, Spain. The ants were kept in artificial nests in Sanyo climate chambers set to 14 hours of light in 27 °C and 10 hours of dark in 21 °C. The artificial nest was a plastic box with a plastered ground in which there were nine breeding boxes with six chambers, also with a plastered ground. The plaster ground was moistened once a week with tap water. There were several dozens of queens in a box and they were kept with workers (several hundreds to thousands, i.e. a large colony) and brood. The ants were fed with honey and cockroaches. The honey was provided from a plastic weighing boat and covered with a piece of paper towel. The cockroach species was commercially bred *Nauphoeta cinerea*, and the adult cockroaches were freeze-killed before feeding to the ants.

The samples, which were reproducing queens in the artificial nests, were initially used for investigating the effects of bacterial injection and social isolation on Argentine ant gene
expression (unpublished). For this study, we confirmed that neither the bacterial injection nor the social isolation affected the virus loads of the queens (bacteria: two-tailed t-test: t = -0.29215, df = 36.403, p-value = 0.7718; social isolation: two-tailed t-test: t = 0.052319 df = 41.554, p-value = 0.9585). The raw data is deposited at NCBI, BioProject PRJNA279338.

The March 2018 samples—40 queens and 40 workers—were collected from the same site as the 2011 samples and preserved in RNAlater (ThermoFisher) until RNA isolation.

RNA isolation

For RNA sequencing the ants were frozen in liquid nitrogen and kept in -80 °C freezer until RNA extraction. The whole-body samples were disrupted and homogenized in a TissueLyser II (Qiagen) using stainless steel beads (5 mm diameter). Total RNA was extracted using an RNeasy Micro Kit (Qiagen), following the protocol provided with the kit. RNA was quantified using an Agilent 2100 Bioanalyzer and the samples sent to BGI Tech Solutions (Hong Kong) for poly-A-mRNA library preparation (Illumina TruSeq RNA Sample Prep Kit) and sequencing (100 bp paired-end reads) with an Illumina HiSeq2000.

The March 2018 samples were isolated with Trizol (ThermoFisher) as recommended by the manufacturer after the ants were frozen with liquid nitrogen and homogenised with a microtube grinder.

Bioinformatics

For virus genome assembly, we used reads that did not map to the _L. humile_ reference genome (GCF_000217595.1). The unmapped sequences from queens were assembled with Trinity v2.3.2 [53] using read normalisation. The assembled contigs were searched for virus protein sequences using NBCI virus protein RefSeq (downloaded January 2018) with
Blastx v2.3.0 [54] with an e-value limit of 0.0001 and a culling limit of 1. The virus sequences were deposited at NCBI GenBank (Table 1 and Table A1). For individual virus load analyses we removed the poly-A-tails from the virus sequences and then mapped only reads that did not map to the *L. humile* reference genome using BWA-MEM [55]. We counted only reads that had a properly paired mate read also mapping to the virus sequence and standardised the virus read amounts by dividing them by the amount of genome mapped reads.

TBLastn was used to identify *L. humile* transcripts from the transcriptome shotgun assembly (TSA) database. Domain enhanced lookup time accelerated Blast was used to identify possible domains and identical proteins sequences. TMHMM Server v. 2.0 [56] was used to predict transmembrane domains. HHPred [57] was used to search for remote homologous proteins based on 3D protein structure prediction. NOVOplasty [58] was used together with pooled genome unmapped *L. humile* DNA sequence data (PRJNA279338) to assemble the LhuEV1 genome sequence. NOVOplasty with CAP3 [59] was used to construct complete virus sequences from the NCBI’s SRA data.

Phylogenies were reconstructed from the RdRp amino acid sequences of each virus, which were aligned using the E-INS-I method in MAFFT v7.313 [60]. The alignments contained a large number of gaps and therefore, prior to phylogenetic analysis, the alignments were trimmed using trimAl v1.2 [61]. The amino acid substitution model was selected using ProtTest 3 [62] and phylogeny reconstruction was performed using PhyML v.3.0 [63].
Substitutions between the 2012 and 2014 virus strains were determined by aligning the most conserved 2012 and 2014 virus strain nucleotide sequences in MEGA 6 [64]. The \( d_N \), \( d_S \) and \( d_N/d_S \) ratios were estimated using the yn00 program in PAML 4 package [65].

**Statistical analysis**

All statistical analyses were performed in R v3.3.3. Expected numbers of queens with different virus combinations were calculated using virus probabilities (no. queens with the virus / all queens) for all possible virus combinations. A queen could have any combination of the seven different viruses (KBV was disregarded as it was in all queens). The expected number of queens with a certain virus combination was thus calculated by multiplying the individual probability of each of the eight viruses being present or not in a queen and multiplying it with the amount of all queens. Expected and observed values for each possible combination were compared with Fisher exact test with a simulated p-value. Pairwise analysis of virus reads was done with Spearman’s correlation using multiple testing p-value correction.

**PCR**

LhuPcV1 replication was tested from the 2014 queen samples by transcribing separately the negative- and positive-sense virus genomes using RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher) using primers 5′ TATAGGTAACCTTTCATGGAAAC and 5′ AGGAGTCAAATTATCAAACTCC, respectively. PCR was made with LhuPcV1 specific primers 5′ TTTGCCTTTCTTAGCTCAGTCC and 5′ CCTCGATGGAGGTATTTTGAAG using FIREpol® (SolisBiodyne). PCR products were gel isolated with a GeneJET gel extraction kit (ThermoFisher) and ligated into the pJET-plasmid with a CloneJET PCR cloning kit (ThermoFisher) and transformed to DHalpha5 cell lines (ThermoFisher). Plasmids were isolated with a GeneJET plasmid mini prep kit (ThermoFisher) and used as
standards for quantitative PCR with EvaGreen (SolisBiodyne), as recommended by the manufacturer. Results were analysed with LightCycler® 96 system software.

Detection of the viruses from the 2018 collected samples was done by producing cDNA with a RevertAid H Minus First Strand cDNA Synthesis Kit (ThermoFisher) using random primers as recommended by the manufacturer. PCR was done with EvaGreen, using the above mentioned primers for LhuPcV1 and the following primers for the rest of the viruses:

- **KBV** 5′ GGTGAAATTGCAAGCCGGT and 5′ TCTGGTTTCCGGGTTGGAC; LhuCV1
- 5′TTGAATGTGGGCAGAGGT and 5′ GGTACCGCAAGATGGAGG; LhuBLV1
- 5′GTCAACAGAGGACGAGATCA and 5′ GCCATAACCTTGGCTCTGC; LhuRLV1: 5′
- AAGGTGAACATGGAAGCTAAT and 5′ AAAATTATCCAGCCCCATTCTCT; LhuQLV1 5′
- TACAGGACAGGGTGCAGTG and 5′ CCCAAGTTGTGAGCAGCAGA; LhuPLV1 5′
- TAGGTTATCAGACCGCGG and 5′ TCGCTAACAAATCGCCGTT; LhuTLV1 5′
- CTTGTCGATATCCGCTCCG and 5′ TCCTGAATTCTAGCAGGCCC; LhuEV1 5′
- CGATAAATCTCCTCCTCTATTGAT and 5′ GACTCGCTAGCAATTCTCTATCA.

**Author statements**

**Funding information**

This research was funded by Academy of Finland grant no. 260147 to LV and Kone Foundation to JJ. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**Acknowledgements**

We would like to thank CSC – IT Center for Science, Finland, for computational resources and the reviewers for their constructive comments.

**Conflicts of interest**

Authors have no conflicting interests to declare relevant to this study.
References


32. Shen M, Yang X, Cox-Foster D, Cui L. The role of varroa mites in infections of Kashmir bee virus (KBV) and deformed wing virus (DWV) in honey bees. *Virology* 2005;342:141–149.


Fig. legends

Fig. 1. Genomic organisation of the RNA viruses discovered from the European Main Argentine ant supercolony. The viruses are grouped according to the genome type: positive-sense, negative-sense, and double-stranded RNA viruses. Line denotes virus sequence, boxes denote open reading frames (ORFs), and the alignment of the boxes on the line indicate different reading frames. Light grey boxes are ORFs with predicted function and dark grey boxes are without. The size in nucleotides is given above each
genome. RdRp-complex denotes all the proteins the virus has for processing RNA in a complex with the RNA-dependent RNA-polymerase.

**Fig. 2.** The phylogenetic association of the RNA viruses discovered from the European Main Argentine ant supercolony to other invertebrate viruses. Phylogenetic trees are shown for A.) LhuCV1, B.) LhuRLV1, C.) LhuBLV1, D.) LhuQLV1, E.) LhuPLV1, and F.) LhuTLV1.

**Fig. 3.** The RNA virus amounts, loads, and interactions in the Argentine ant queens. A.) Individual standardised read amounts for the eight RNA viruses were blotted for each queen (N=48). Virus colour code and the number of the queens that have the virus are given under the scatter plot. B.) Each virus has its typical load and activation level. Variation of all the virus reads (ALL) and virus-specific reads in the queens are shown by box plots. Black line represents median value, values inside the box cover first and third quartile (interquartile) range, the whisker shows values up to 1.5 times the interquartile range, and grey circles are individual values beyond the whisker range (outliers). C.) The queens were infected on average by three to four different RNA viruses. The amount of different virus types per queen is shown in the bar plot and the colours indicate the viruses that produced the combinations. For example, the three queens that were infected with six viruses all had the same virus combination of KBV, LhuPcV1, LhuBLV1, LhuRLV1, LhuPLV1, and LhuTLV1. Virus colour codes are given under the bar plot. D.) Pairwise comparison of virus read amounts was analysed by Spearman’s correlation coefficient. Positive correlation is indicated by increasing darkness of blue, negative correlation by red, and no correlation by white, as indicated by the colour scale. Multiple test adjusted p-values (q-values) are presented for each comparison. The strength of q-values is indicated
by the size of the coloured circle. The only significant result is the negative correlation between LhuBLV1 and LhuQLV1, which were never found to infect the same queen.

**Fig. 4.** Interaction map of the Argentine ant RNA viruses. Original Argentine ant viruses are presented as coloured squares and potentially new viruses as circles. The thickness of the connecting line illustrates the number of queens the viruses both infect. For example, KBV and LhuPcV1 infect the same 41 queens whereas LhuPcV1 and LhuTLV1 infect six. Black connection lines indicate the two most common combinations of viruses: i) KBV, LhuPcV1, and LhuPLV1 and ii) KBV, LhuPcV1, LhuBLV1, LhuPLV1, and LhuCV1. The red dashed line shows the number of expected interactions between LhuBLV1 and LhuQLV1 not found from our samples.

**Table 1. The presence of the viruses found from our study in Argentine ant supercolonies and other insects.** The genome type of the virus is shown in the parenthesis after the virus name. EM = European Main, NZ = New Zealand, and CA = Californian supercolony. SRA = sequence read archive, TSA = Transcriptome Shotgun Assembly database.

**Table 2. The genome evolution and infection activity values of the Argentine ant RNA viruses.** The genome type of the virus is shown in parentheses after the virus name. $d_N$ = non-synonymous substitutions, $d_S$ = synonymous substitutions.

**Appendixes**

**Fig. A1.** The genomic structure of the new picorna-like (LhuPiLV1), narna-like (LhuNLV1), and polycipi (LhuPcV2) viruses found from *L. humile* SRA data. The mRNA contigs from Bioprojects PRJDB4088 and PRJEB10079 were analysed with Blastx against the NCBI...
virus protein database and complete virus contigs were identified. LhuPiLV1 RdRp is, according to Blastp, 47 % identical with 98 % coverage to Hubei picorna-like virus 24 hypothetical protein (YP_009329985.1) and 37 % identical with 80 % coverage to nonstructural polyprotein of Black queen cell virus (APZ86805.1). LhuPiLV1 capsid is 41 % identical with 60 % coverage to hypothetical protein 2 of Hubei picorna-like virus 24 (YP_009329986.1). LhuNLV1 has two potential ORFs, coding on both sense and antisense directions. LhuNLV1 RdRp is 34 % identical with 95 % coverage to RdRp of Hubei narna-like virus 19 (YP_009337792.1). Hypothetical LhuNLV1 ORF1 is 29 % identical with 37 % coverage to L. niger hypothetical protein RF55_14147 (KMQ86783.1) and 31 % identical with 23 % coverage to hypothetical protein of Hubei narna-like virus 20 (APG77160.1). LhuPcV2 has five putative ORFs, from which ORF1, 3, and 4 code, according to Blasp, proteins similar to putative capsid proteins (ORF1: 40 % identical with 94 % coverage to SINV2 putative structural protein (YP_001285726.1), ORF3: 40 % identical with 97% coverage to SINV2 putative structural protein (YP_001285727.1), and ORF4: 46 % identical with 65 % coverage to SINV4 putative capsid protein (YP_009407936.1). According to HHpred and Blastp ORF4b have no significant similarity to any known proteins and ORF2 is similar to hypothetical protein of Lasius niger virus 1 (YP_009407939.1). ORF2b is according to TMHMM a transmembrane protein, which is typical to ant polycipiviruses. ORF5 codes for RdRp-complex, which is 41 % identical with 97 % coverage to Lasius niger virus 1 (YP_009407943.1). The line denotes virus sequence, boxes denote open reading frames (ORFs), and the alignment of the boxes on the line indicate different reading frames. Light grey boxes are ORFs with predicted function and dark grey boxes without. The sequence size in nucleotides is given above each genome.
Fig. A2. Phylogenetic association of the new picorna-like (LhuPiLV1), narna-like (LhuNLV1), and polycipi (LhuPcV2) viruses found from L. humile SRA data. A.) LhuPiLV1 is most related to Hubei picorna-like virus 24 (ID: 51278) isolated from myriapoda mix and also to other picorna-like viruses, such as SINV5 of Solenopsis invicta and honey bee pathogen BQCV. B.) LhuNLV1 is related to other narna-like viruses and most closely to Hubei narna-like virus 19 (ID: 52343) isolated from odonata mix. C.) LhuPcV2 is similar to other polycipivirus, such as LhuPcV1, Lasius niger virus 1, and Solenopsis invicta virus 2.

Fig. A3. LhuPcV1 replicates in the Argentine ant queens. cDNA specific for the replicative form (negative-sense, grey bars) and genome/mRNA (positive-sense, orange bars) of LhuPcV1 were produced from five queens and used as a qPCR template with LhuPcV1 specific primers.

Table A1. The presence of all of the complete viruses identified in this study in other Argentine ant supercolonies based on SRA data search. The virus name is shown in parentheses after the genome type of the virus. EM = European Main, NZ = New Zealand, CA= Californian supercolony.
Fig. 2.

a. Diaphorina citri associated C virus
   Wuhan insect virus 35
   Jingmen tombus-like virus 2
   Hubei tombus-like virus 38
   Wenzhou crab virus 4
   Sansia tombus-like virus 9
   Hubei tombus-like virus 40
   Chronic bee paralysis virus
   Linopithema humile C-virus 1
   Antaphelone associated C virus

b. Beihai rhabdovirus 5
   Beihai rhabdovirus 4
   Hubei rhabdovirus 1
   Hubei rhabdovirus 2
   Hubei rhabdovirus 9
   Culex rhabdovirus
   Hubei rhabdovirus 4
   Hubei rhabdovirus 3
   Linopithema humile rhabdovirus 1
   Beihai rhabdovirus 2
   Beihai rhabdovirus 1
   Hubei rhabdovirus 7

c. Cumulo virus
   Shabe heteroptera virus 3
   Zhe Mosquito virus
   Hubei bunya-like virus 3
   Linopithema humile bunya-like virus 1
   Wuhan insect virus 16
   Wenzhou Shrimp Virus 1
   Huanggish Tick Virus 2
   Yongji Tick Virus 1
   Wuhan crustacean virus 7

d. Linopithema humile qimivirus-like virus 1
   Wuhan insect virus 15
   Hubei qimivirus-like virus 1
   Wenzhou qimivirus-like virus 1
   Linopithema humile qimivirus-like virus 2
   Wuhan insect virus 16
   Wenzhou Shrimp Virus 1
   Huanggish Tick Virus 2
   Yongji Tick Virus 1
   Wuhan crustacean virus 7
   Shabe qimivirus-like 1

e. Hubei partili-like virus 17
   Wuhan insect virus 22
   Hubei partili-like virus 22
   Wuhan partili-like virus 3
   Wuhan partili-like virus 2
   Hubei partili-like virus 10
   Wuhan cricket virus 2
   Hubei partili-like virus 2
   Linopithema humile partili-like virus 1
   Hubei phtera virus 17

f. Hubei toti-like virus 17
   Wenzhou crab virus 5
   Beihai paphia shell virus 5
   Hubei toti-like virus 16
   Beihai toti-like virus 5
   Hubei toti-like virus 23
   Hubei toti-like virus 19
   Sansia water strider virus 20
   Linopithema humile toti-like virus 1
   Shuangao toti-like virus
Fig. 3.

(a) Scatter plot showing the number of virus genome mapped reads per queen. Each point represents a queen, with the color indicating the type of virus and the size indicating the number of reads.

(b) Box plot showing the distribution of virus genome mapped reads for different RNA virus types. The box plot includes the median, interquartile range, and outliers for each type.

(c) Bar chart showing the number of queens for different RNA virus types per queen. The bar color indicates the type of virus.

(d) Heatmap showing the correlation between RNA virus types and queen number. The color scale represents the correlation coefficient.
Fig. 4.
Table 1. The presence of the viruses found from our study in Argentine ant supercolonies and other insects.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Genbank ID</th>
<th>2012 EM TSA ID</th>
<th>2018 EM (N=40)</th>
<th>In NZ (SRA)</th>
<th>In CA (SRA)</th>
<th>In other insects (TSA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KBV (+)</td>
<td>LI706695.1</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td></td>
<td>Yes, Honey bees (Apis mellifera) and ants (Pogonomyrmex californicus, Formica exsecta, Formica pressilabris, Formica fusca, Formica cinerea)</td>
</tr>
<tr>
<td>LhuPcV1 (+)</td>
<td>MH213247</td>
<td>LI719284.1</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LHuCV1 (+)</td>
<td>MH213244 MH213245</td>
<td>Li682236.1 (Capsid) Li681900.1 (RdRp)</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuBLV1 (-)</td>
<td>MH213247 MH213248</td>
<td>LI727600.1 (RdRp) LI712600.1 (Capsid)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>LhuRLV1 (-)</td>
<td>MH213246</td>
<td>LI702938.1 (ORF1-2) LI691951.1 (ORF3) LI681947.1 (ORF4-5) LI722247.1 (RdRp)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuQLV1 (-)</td>
<td>MH213241 MH213242</td>
<td>LI762419.1 (Capsid) LI719217.1 (RdRp)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>Similar virus in ant Crematogaster osakensis</td>
</tr>
<tr>
<td>LhuPLV1 (ds)</td>
<td>MH213239 MH213240</td>
<td>LI704840.1 (Capsid) LI696689.1 (RdRp)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuTLV1 (ds)</td>
<td>MH213243</td>
<td>LI706803.1 (27% coverage)</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuEV1 (dsDNA)</td>
<td>MH213250</td>
<td>LI761971.1 (2 genes)</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>NA</td>
</tr>
</tbody>
</table>

The genome type of the virus is shown in the parenthesis after the virus name. EM = European Main, NZ = New Zealand, and CA = Californian supercolony. SRA = sequence read archive, TSA = Transcriptome Shotgun Assembly database.
Table 2. The genome evolution and infection activity values of the Argentine ant RNA viruses.

<table>
<thead>
<tr>
<th>Virus</th>
<th>alignment length</th>
<th>substitutions</th>
<th>d_N</th>
<th>d_S</th>
<th>d_N/d_S</th>
<th>prevalence (%)</th>
<th>no. high outliers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KBV (+)</td>
<td>7935</td>
<td>98</td>
<td>0.0032</td>
<td>0.0414</td>
<td>0.078</td>
<td>100</td>
<td>8 (16.7)</td>
</tr>
<tr>
<td>LhuPcV1 (+)</td>
<td>11199</td>
<td>424</td>
<td>0.0034</td>
<td>0.1685</td>
<td>0.020</td>
<td>85.4</td>
<td>8 (19.5)</td>
</tr>
<tr>
<td>LHuCV1 (+)</td>
<td>5830</td>
<td>57</td>
<td>0.0035</td>
<td>0.0213</td>
<td>0.163</td>
<td>25.0</td>
<td>0</td>
</tr>
<tr>
<td>LhuRLV1 (-)</td>
<td>11908</td>
<td>11</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>43.8</td>
<td>2 (9.5)</td>
</tr>
<tr>
<td>LhuBLV1 (-)</td>
<td>6725</td>
<td>3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>54.2</td>
<td>0</td>
</tr>
<tr>
<td>LhuQLV1 (-)</td>
<td>6628</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>16.7</td>
<td>0</td>
</tr>
<tr>
<td>LhuPLV1 (ds)</td>
<td>2902</td>
<td>4</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>41.7</td>
<td>1 (5.0)</td>
</tr>
<tr>
<td>LhuTVL1 (ds)</td>
<td>1933</td>
<td>3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>12.5</td>
<td>1 (16.7)</td>
</tr>
</tbody>
</table>

The genome type of the virus is shown in the parenthesis after the virus name. d_N = non-synonymous substitutions, d_S = synonymous substitutions.
**Supplementary data**

**Fig. S1.** The genomic structure of the new picorna-like (LhuPiLV1), narna-like (LhuNLV1), and polycipi (LhuPcV2) viruses found from *L. humile* SRA data. The mRNA contigs from Bioprojects PRJDB4088 and PRJEB10079 were analysed with Blastx against NCBI virus protein database and complete virus contigs were identified. LhuPiLV1 RdRp is according to Blastp 47 % identical with 98 % coverage to Hubei picorna-like virus 24 hypothetical protein (YP_009329985.1) and 37 % identical with 80 % coverage to nonstructural polyprotein of Black queen cell virus (APZ86805.1). LhuPiLV1 capsid is 41 % identical with 60 % coverage to hypothetical protein 2 of Hubei picorna-like virus 24 (YP_009329986.1). LhuNLV1 has two potential ORFs, coding on both sense and antisense directions. LhuNLV1 RdRp is 34 % identical with 95 % coverage to RdRp of Hubei narna-like virus 19 (YP_009337792.1). Hypothetical LhuNLV1 ORF1 is 29 % identical with 37 % coverage to *L. niger* hypothetical protein RF55_14147 (KMQ86783.1) and 31 % identical with 23 % coverage to hypothetical protein of Hubei narna-like virus 20 (APG77160.1). LhuPcV2 has
five putative ORFs, from which ORF1, 3, and 4 code according to Blasp proteins similar to putative capsid proteins (ORF1: 40 % identical with 94 % coverage to SINV2 putative structural protein (YP_001285726.1), ORF3: 40 % identical with 97% coverage to SINV2 putative structural protein (YP_001285727.1), and ORF4: 46 % identical with 65 % coverage to SINV4 putative capsid protein (YP_009407936.1). According to HHpred and Blastp ORF4b have no significant similarity to any known proteins and ORF2 is similar to hypothetical protein of Lasius niger virus 1 (YP_009407939.1). ORF2b is according to TMHMM a transmembrane protein, which is typical to ant polycipiviruses. ORF5 codes for RdRp-complex, which is 41 % identical with 97 % coverage to Lasius niger virus 1 (YP_009407943.1). Line denotes for virus sequence, boxes open reading frames (ORFs), and the alignment of the boxes on the line indicate different reading frames. Light grey boxes are ORFs with and dark grey boxes without predicted function. The sequence size in nucleotides is given above each genome.
**Fig. S2.** Phylogenetic association of the new picorna-like (LhuPiLV1), narna-like (LhuNLV1), and polycipi (LhuPcV2) viruses found from L. humile SRA data. A.) LhuPiLV1 is most related to Hubei picorna-like virus 24 (ID: 51278) isolated from myriapoda mix and also to other picorna-like viruses such as SINV5 of Solenopsis invicta and honey bee pathogen BQCV. B.) LhuNLV1 is related to other narna-like viruses and most closely to Hubei narna-like virus 19 (ID: 52343).
isolated from odonata mix. C.) LhuPcV2 is similar to other polycipiviruses such as LhuPcV1, Lasius niger virus 1, and Solenopsis invicta virus 2.
**Fig. S3.** LhuPcV1 replicates in the Argentine ant queens. CDNA specific for the replicative form (negative-sense, grey bars) and genome/mRNA (positive-sense, orange bars) of LhuPcV1 were produced from five queens and used as a qPCR template with LhuPcV1 specific primers.
Table S1. The presence of all of the complete viruses identified in this study in other Argentine ant supercolonies based on SRA data search.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Genbank ID</th>
<th>In EM colony</th>
<th>In NZ colony</th>
<th>In CA colony</th>
</tr>
</thead>
<tbody>
<tr>
<td>LhuPcV1 (+)</td>
<td>MH213247</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuPcV2 (+)</td>
<td>MH213248</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>LHuCV1 (+)</td>
<td>MH213244, MH213245</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LHUV1 (+)</td>
<td>MH213249</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>LhuPILV1 (+)</td>
<td>MH213235</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuNLV1 (+)</td>
<td>MH213236</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuBLV1 (-)</td>
<td>MH213247, MH213248</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>LhuRLV1 (-)</td>
<td>MH213246</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>LhuQLV1 (-)</td>
<td>MH213241, MH213242</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>LhuPLV1 (ds)</td>
<td>MH213239, MH213240</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>LhuTLV1 (ds)</td>
<td>MH213243</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

The genome type of the virus is shown in the parenthesis after the virus name. EM = European Main, NZ = New Zealand, CA = Californian supercolony.