African army ants at the forefront of virome surveillance in a remote tropical forest

Matthieu Fritz[1*], Bérénice Reggiardo[2,3*], Denis Filloux[2,3], Lisa Claude[2,3], Emmanuel Fernandez[2,3], Frédéric Mahé[2,3], Simona Kraberger[4], Joy M. Custer[4], Pierre Becquart[1], Telstar Ndong Mebaley[5], Linda Bohou Kombila[5], Léadisaelle H. Lenguiya[6,7], Larson Boundenga[5,8], Illich M. Mombo[5], Gael D. Maganga[5,9], Fabien R. Niama[6,7], Jean-Sylvain Koumba[5], Mylène Ogliastro[10], Michel Yvon[3], Darren P. Martin[11], Stéphane Blanc[3], Arvind Varsani[4,12], Eric Leroy[1*]# and Philippe Roumagnac[2,3*]#

[1] Institut de Recherche pour le Développement (IRD), Maladies Infectieuses et Vecteurs, Ecologie, Génétique, Evolution et Contrôle (MIVEGEC) (Université de Montpellier-IRD 224–CNRS 5290), 34394 Montpellier, France

[2] CIRAD, UMR PHIM, 34090 Montpellier, France

[3] PHIM Plant Health Institute, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France

[4] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA

[5] Centre Interdisciplinaire de Recherches Médicales de Franceville, Franceville, Gabon

[6] Laboratoire National de Santé Publique, Brazzaville, Republic of Congo

[7] Université Marien Ngouabi, Brazzaville, Republic of Congo

[8] Department of Anthropology, Durham University, South Road, Durham DH1 3LE, UK.

[9] Université des Sciences et Technique de Masuku (USTM), Institut National Supérieur d'Agronomie et de Biotechnologies (INSAB), Franceville, Gabon

[10] INRA-Université de Montpellier UMR DGIMI 34095 Montpellier, France

[11] Division of Computational Biology, Department of Integrative Biomedical Sciences, Institute of infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa

[12] Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town, South Africa

* These authors contributed equally to this work.

For correspondence. E-mail: philippe.roumagnac@cirad.fr and eric.leroy@ird.fr

Tel. +33 (0)4 99 62 48 53

Abstract

In this study, we used a predator-enabled metagenomics strategy to sample the virome of a remote and difficult-to-access densely forested African tropical region. Specifically, we focused our study on the use of army ants of the genus *Dorylus* that are obligate collective foragers and group predators that attack and overwhelm a broad array of animal prey. Using 209 army ant samples collected from 29 colonies and the virion-associated nucleic acid-based metagenomics approach, we showed that a broad diversity of bacterial, plant, invertebrate and vertebrate viral sequences were accumulated by army ants: including sequences from 157 different viral genera in 56 viral families. This suggests that using predators and scavengers such as army ants to sample broad swathes of tropical forest viromes can shed light on the

composition and the structure of viral populations of these complex and inaccessible ecosystems.

Introduction

Viruses are likely the most abundant and diverse biological entities on Earth [1-3] and are arguably the most successful inhabitants of the biosphere [4]. Despite this, the current inventory of known virus diversity is likely a vanishingly small and unrepresentative fraction of the total diversity, abundance, and population structures of all extant viruses [1,5-7]. Hence, while the overall total number of eukaryotic virus species on Earth is estimated to exceed 87 million [1], only 10,434 virus species (including eukaryotic and prokaryotic viruses) are presently recognized by the International Committee on Taxonomy of Viruses [8]. Our perception of the true extent and properties of the virosphere is further clouded by the fact that this minuscule sample of global virus diversity is heavily biased towards virus species that directly impact humans and the organisms that we farm [6,7].

The best studied components of the virosphere are those that include the plant, animal, fungi and bacteria-infecting viral agents – called viromes – within or upon human bodies, human food sources (especially domesticated plants and animals) and human habitats (especially urban homes, hospitals, schools and farms). Conversely, the least studied components of the virosphere are those of natural environments, particularly remote ecosystems such as those found in deep tropical forests. Crucially, even when metagenomic projects have explored the viromes in such regions, these studies have been geographically and taxonomically biased. Hence, samples were only accessible via forest roads or tracks, and derived from the subset of plant or animal species that are likely to host viruses with some medical or agricultural relevance to humans [9]. Such sampling biases are understandable in that most future viral

diseases of humans and their domesticated plants and animals will likely emerge from host species living or wandering in their vicinity. In addition, future emerging diseases will often be related to viruses already known to cause diseases in humans, and domesticated plants and animals. However, it is important to keep in mind that a less human-centric assessment of viral diversity at the ecosystem-scale could i) illuminate the natural host-ranges, ecological contexts and evolutionary processes underlying the diversification of viruses related to those that have already emerged to cause diseases that impact humans, and ii) identify the yet unknown viruses with properties such as broad host ranges or high incidences that could potentially pose future threats to humans [9,10]. More generally, the paucity of information on virus diversity in natural environments is hampering our understanding of both the roles of viruses within wild ecosystems, and how natural- or human- mediated disturbances of such ecosystems might impact these roles [10-12].

Densely forested tropical regions account for 40 % of the world's 4 billion hectares of forests [13] and represent major interfaces where humans interact with the world's remaining wilderness areas. Besides accessibility issues, studying viromes in these habitats poses several additional challenges. The spatial pattern of sampling schemes, the establishment of a list of organisms to sample, and the developmental stages and/or symptom statuses of these organisms are all important considerations as they may or may not yield metagenomic data representative of the diversity of viruses circulating within the targeted area [14]. An interesting "meta-sampling" strategy that provides an alternative to tentative random/unbiased sampling in inaccessible sampling sites is to rely on proxy samplers such as highly mobile predator/scavenger animals that naturally accumulate animal-, fungal- and plant-derived biomass within their digestive tracts during feeding [15-19]. Rather than aiming for completely random sampling of biological material within a given environment, the general intention of this approach, -which is variously

referred to as "xenosurveillance" or "vector-enabled metagenomics" (VEM) - is to shift the causes of unrepresentative sampling from human choice biases to the feeding-choice biases of the chosen mobile scavenger/predator that is intimately involved with the targeted environments [15-19]. While feeding-choice biases are unavoidable with such meta-sampling schemes, they are biologically meaningful in that viral cargos of blood-feeding, sap-feeding and insectivorous insects are themselves inherently representative of viral mobility within ecosystems. Insectivorous arthropods, so-called top-end insect predators, are potentially particularly useful samplers in this regard. Indeed, the diversity of viruses that they ingest via their prey should be partly representative of the viruses that are ferried within insects throughout the ecosystem. In addition, sampling top-end insect predators will avoid the capture (and sometimes sacrifice) of animals that are protected for ethical, legal or cultural reasons. Among xenosurveillance approaches, Predator Enabled Metagenomics (PEM), using for example dragonflies[20] or damselflies, [21] has proved more efficient with respect to collecting a wide diversity of viral sequences than Vector-Enabled Metagenomics (VEM) using exclusively plant sap-feeding insects (e.g. whiteflies [19]) or blood-feeding insects (e.g. mosquitoes [15-17]).

In this study, we propose a PEM strategy using army ants of the genus *Dorylus* [22,23] as top-end insect predators. These nomadic insects are obligate collective foragers and group predators that attack a broad array of animal prey including crickets, cockroaches, earthworms, and even vertebrates. Although nomadic, they do have temporary nests around which they daily forage areas of ~1km ^{2} [22-24]. Army ants generally live in colonies containing between 10^4 to 10^7 workers [25] with a daily colony-wide intake of prey/scavenged biomass of up to 2 kg [26]. In addition to preying on an extremely diverse array of animal species, army ants also scavenge on the carcasses of large vertebrates and feed directly on plants [24,27,28].

Here we test the overarching hypothesis that army ants hunting live invertebrate and vertebrate prey in the deep forest ingest and accumulate a diverse array of plant and animal viruses in the areas around their temporary nests.

Materials and Methods

Army ants sampling

In July 2019, two sampling surveys were conducted in the Ogooué-Ivindo region (Northeast Gabon). Over 250 individual ants from 29 colonies were collected along roadsides between Mekambo and Mendemba villages, immediately stored in liquid nitrogen and further transferred to Montpellier (France) where they were stored at -80°C. Two-hundred and nine samples, including 145 samples containing only one ant and 64 samples each containing pools of 2 to 13 individual ants (Table 1 and Supplementary Table 1) were further processed using the virion-associated nucleic acid based (VANA) metagenomics approach [29].

Virion-associated nucleic acid-based viral metagenomics

Each of the 209 collected samples was processed using the VANA viral metagenomics approach that is comprehensively detailed in François et al. (2018) [30]. It is noteworthy that the VANA approach is suited to the detection of both DNA and RNA viruses, and includes several steps that aim at both removing host nucleic acids and maximizing the yield of virion-associated nucleic acids [29]. Briefly, ants were first ground using tissue homogenizer and sterile steel beads. Viral particles from individual or pooled ant samples were first isolated using centrifugation and filtration techniques and they were further concentrated by ultracentrifugation. Contaminating non-encapsidated nucleic acids were then digested by DNase and RNase digestion treatments. Following this, encapsidated DNA and RNA molecules resistant to the DNase and RNase treatments were extracted. A series of molecular amplification was subsequently carried out, including reverse

Table 1: Characteristics of the army ant samples. P21/P23 means that colonies P21 and P23 were pooled. The species assignation is based on Cytochrome oxidase I reads and contigs analysis. "nd" means non determined.

Ant colony number	Nb. of samples	Genus	Species	Foraging niche	Collection date	Collection location	GPS coordinates
FM1	4	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-07	Mendemba	nd
FM2	4	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-07	Mendemba	nd
FM3	4	*Dorylus* sp.	nd	nd	2019-07-07	Between Ntolo and Mekouma	nd
FM4	4	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-07	Between Ntolo and Mekouma	nd
FM5	4	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-07	Mekouma	nd
FM6	4	*Dorylus* sp.	nd	nd	2019-07-07	Between Mekouma and Malassa	nd
FM7	4	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-07	Mekambo	nd
FM8	4	*Dorylus* sp.	nd	nd	2019-07-07	Mekambo	nd
P1A	9	*Dorylus* sp.	nd	nd	2019-07-08	Mekambo	00°57.214' - 013°58.396'
P2A	5	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-08	Komambela	00°49.865' - 013°58.846'
P3A	8	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-09	Mekambo	00°56.681' - 013°59.032'
P4A	8	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-09	Between Ibong and Zoula	01°02.872' - 014°00.951'
P5	8	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-09	Between Ibea and Zoula	01°04.409' - 014°03.100'
P6	8	*Dorylus* sp.	*D. opacus/kohli* clade	leaf-litter	2019-07-09	Between Zoula and Grand Etoumbi	01°08.545' - 014°06.826'
P7	8	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-09	Between Massombo and Egopouma	01°16.394' - 014°06.873'
P8A	9	*Dorylus* sp.	nd	nd	2019-07-09	Between Ibea and Zoula	01°04.086' - 014°02.629'
P9	9	*Dorylus* sp.	nd	nd	2019-07-09	Between Ibea and Zoula	01°03.571' - 014°02.068'
P10	9	*Dorylus* sp.	nd	nd	2019-07-10	Between Mbessa and Malassa	00°54.502' - 014°01.582'
P11	9	*Dorylus* sp.	nd	nd	2019-07-12	Between Malassa and Mekouma	00°51.149' - 014°04.198'
P12	9	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-12	Between Ntolo and Mekouma	00°49.975' - 014°04.212'
P13	9	*Dorylus* sp.	nd	nd	2019-07-10	Between Mekouma and Ntolo	00°48.731' - 014°04.110'
P14	9	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-10	Between Mekouma and Ntolo	00°44.720' - 014°05.251'
P15	9	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-11	Mekambo	00°59.282' - 013°57.916'
P16	9	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-11	Mekambo	00°58.182' - 013°58.277'
P17	9	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-11	Mekambo	00°57.042' - 013°58.532'
P18	9	*Dorylus* sp.	nd	nd	2019-07-11	Mbeza	00°55.154' - 013°59.798'
P19	9	*Dorylus* sp.	nd	nd	2019-07-12	Between Mekambo and Mbeza	00°56.939' - 013°58.605'
P21/P23	5	*Dorylus* sp.	*D. sjoestedti/wilverthi* clade	surface swarms	2019-07-12	Mbeza	00°55.167' - 013°59.793' / 00°52.804' - 014°03.581'
P24	10	*Dorylus* sp.	nd	nd	2019-07-12	Between Malassa and Mekouma	00°52.253' - 014°03.609'

transcription, Klenow fragment treatment, and amplification of the viral DNA and RNA using barcoded PCR primers. Finally, amplification products were pooled into 3 libraries (each composed of 72 amplification products including army ant samples and positive/negative controls) and sequenced by Genewiz (Leipzig, Germany) using a single lane on an Illumina HiSeq 3000/HiSeq 4000 sequencer (2 × 150 bp sequencing). Bioinformatics analyses were performed as described previously [30]. Briefly, demultiplexing was performed with the agrep command-line tool to assign reads to the samples from which they originated [31]. Adaptors were removed and the reads were filtered for quality (q30 quality and read length >45 nt) using Cutadapt 3.1 [32]. The cleaned reads were assembled *de novo* into contigs using SPAdes 3.6.2 [33]. Putative virus reads obtained using BLASTx [34] with e-values < 0.001 were retained. Amplification products (amplicons) of four samples (N° 166, 185, P13-3.2 and P17-2.2, Supplementary Table 1) were also sequenced in parallel using the recently developed Flongle (flow cell dongle) sequencing system (Oxford Nanopore Technologies, Oxford, UK). These amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). Library construction for the Flongle sequencing system was performed using the SQK-LSK109 Kit, following the manufacturer's instructions. Four Flongle Flow Cell (R9.4.1) were used for sequencing. The bioinformatics analysis of the Nanopore reads was carried out as follows: accurate base calling was performed using Guppy (v5.0.16; available online at https://nanoporetech.com/). Adapter and primer (Dodeca linker) removal was then performed using Porechop v0.2.4 (available online at https://github.com/rrwick/Porechop). The quality of reads was investigated using NanoPlot v1.33.0 [35]. Taxonomic assignment was achieved on cleaned Nanopore reads through searches against the NCBI nr protein database using DIAMOND 0.9.22 with an e-value threshold of < 0.001 [36].

Dorylus sp. cytochrome oxidase I reads inventory and taxonomic assignment

Illumina reads assigned to *Dorylus* sp. *cytochrome oxidase I* gene using BLASTx searches were recovered and further assembled using SPAdes [33]. Contigs and representative *cytochrome oxidase I* gene sequences of *Dorylus* sp. specimens representing all six recognized subgenera [37] were subsequently aligned using MUSCLE with default settings [38]. A phylogenetic tree was constructed using the maximum likelihood method implemented in PhyML 3.1 [39]. The HKY85 substitution model was selected assuming an estimated proportion of invariant sites of 0.542 and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.219). Support for internal branches was assessed using the aLRT test (SH-Like).

Inventory of virus contigs and putative taxonomic assignment

One of the critical aspects of using BLAST-based searches to assign contigs to viral families is the minimum length of the contig being queried. We have recently conducted a simulation experiment that has revealed that the accuracy of the BLASTx virus family and genus assignations are high, *i.e.* >97.9% and >90.6%, respectively, when contigs have lengths ≥200 nt [29]. We therefore selected contigs with lengths ≥200 nt and retained viral BLASTx assignations of these contigs wherever they yielded e-values < 0.001.

Statistical analyses

The taxonomic assignations of ≥200 nt long virus contigs at the genus level (as determined by best BLASTx hits) were scored for individual ants, including 67 individual workers and 78 individual soldiers. Differences between the number of viral genera associated with workers and soldiers were compared using Mann-Whitney U-tests. Differences between the number of viral genera associated with individual ants belonging to the *Dorylus sjoestedti/wilverthi* clade (n = 76) and to the *Dorylus opacus/kohli* clade (n = 6) were also compared using Mann-Whitney

U-tests. Finally, differences between the number of viral genera associated with individual ants from 20 *Dorylus* colonies ($4 \leq n \leq 8$) were compared using the Kruskal-Wallis H test.

Phylogenetic analyses of most prevalent virus families

Contigs assigned to the families *Bidnaviridae*, *Dicistroviridae*, *Hepeviridae*, *Iflaviridae*, *Microviridae*, *Nodaviridae*, *Picobirnaviridae*, *Polycipiviridae, Tombusviridae, Tymoviridae* and *Solemoviridae*, as well the Picorna-like virus group and the CP-based sequence grouping of viruses called cruciviruses [40], were translated and conserved protein sequences were extracted (the even more prevalent parvo- and cycloviruses are analysed separately, see next sections). These protein sequences together with representative sets of protein sequences belonging to the virus groups/families to which the sequences were taxonomically assigned, were aligned using MUSCLE with default settings [38]. Neighbor joining phylogenetic trees were generated using MEGA version X [41] using alignments of major capsid protein sequences (*Bidnaviridae* and *Microviridae*), coat protein sequences (CP; cruciviruses), RNA-dependent RNA polymerase sequences (RdRp; *Dicistroviridae*, *Nodaviridae*, *Picobirnaviridae* and *Solemoviridae*), polyprotein sequences (*Hepeviridae*, *Iflaviridae*, Picorna-like viruses, *Tombusviridae* and *Tymoviridae*), and ORF5 sequences (*Polycipiviridae*). One thousand bootstrap replicates were performed to quantify branch support.

Phylogenetic analyses of parvovirus-related sequences

We attempted to evaluate the genetic relationships of the parvovirus-related sequences (PRS) that were by far the most abundant virus sequences amplified from the army ant samples. To reconstruct the evolutionary relationships of the various major parvovirus lineages we focused exclusively on the SF3 (Super Family 3) helicase domain of parvovirus protein, NS1. The SF3 helicase domain is highly conserved in all known parvoviruses and is therefore typically used for phylogenetic analyses of divergent parvoviruses [42,43]. NS1 BLASTx assignations (with e-

values < 0.001) were initially retained and translated *in silico* using ORF finder (cut off ≥ 500 bp) (http://www.ncbi.nlm.nih.gov/projects/gorf/). NS1 protein sequences that were ≥200 aa in length were then selected and processed using the software Interproscan which predicted the presence of functional domains [44]. Four-hundred and three SF3 protein sequences were further selected and combined with both 125 representative SF3 protein sequences of viruses from publicly available transcript, genome and protein databases that were classified as belonging to the *Parvoviridae* family [43] and 155 SF3 protein sequences collected from available genomes of viruses classified in the *Parvoviridae* family from the NCBI genome database. The 683 SF3 protein sequences were aligned using MUSCLE with default settings [38]. A Neighbor-Joining tree was produced using MEGA version X [41] with 1000 bootstrap replicates to quantify branch support. Specifically, NS1 protein sequences that were ≥200 aa in length and which were assigned to the genus *Chaphamaparvovirus* were aligned together with 32 representative protein sequences from this genus using MUSCLE with default settings [38]. A phylogenetic tree was constructed using the maximum likelihood method implemented in PhyML 3.1 [39]. The WAG substitution model was selected by PhyML 3.1 assuming an estimated proportion of invariant sites of 0.075 and four gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.337). Support for internal branches was assessed using the aLRT test (SH-Like).

Sequencing and phylogenetic analyses of complete cyclovirus genomes

Four hundred and seventy-two contigs ≥ 200 nt in length that shared similarity with members of the *Circoviridae* family were identified using BLASTx searches with e-values < 0.001 (Supplementary Table 2). These contigs were initially aligned using MUSCLE and clustered in 22 genetic groups (data not shown). Abutting primer pairs were designed that were specific to these genetic groups (Supplementary Table 3) to enable the recovery of full viral genomes using polymerase chain reaction (PCR) with HiFi HotStart DNA polymerase (Kapa Biosystems,

USA) using cycling conditions per the manufacturer's instructions. The amplicons were resolved on a 0.7% agarose gel using electrophoresis, and amplicons approximately 2–3 kb in size were excised from the gel and purified using the Quick-spin PCR Product Purification Kit (iNtRON Biotechnology, Korea). The amplicons were further cloned in pJET1.2 cloning plasmid (ThermoFisher Scientific, USA), and the recombinant plasmids were then transformed into XL1 blue *Escherichia coli* competent cells. The resulting plasmids from the transformants were purified using a DNA-spin Plasmid DNA Purification kit (iNtRON Biotechnology, Korea), and then sequenced by primer walking at Macrogen Inc. (Korea). Specifically, 45 translated Rep sequences recovered from 45 complete genome sequences and representative cyclovirus Rep protein sequences were aligned using MUSCLE with default settings [38]. A phylogenetic tree was constructed using the maximum likelihood method implemented in PhyML 3.1 [39]. The LG+I+G substitution model was selected assuming an estimated proportion of invariant sites of 0.075 and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.337). Support for internal branches was assessed using the aLRT test (SH-Like).

Results and discussion

Classification of army ants using mitochondrial cytochrome oxidase I gene

Before examining the diversity of viruses associated with the army ant samples that we collected from Gabon, we attempted to determine the *Dorylus* species that the ants belonged to by examining their mitochondrial *cytochrome oxidase I* genes. Ninety-nine reads from 24/209 of the ant samples, representing 17/29 of the ant colonies, that shared identity with the *cytochrome oxidase I* gene were recovered from the Illumina sequencing run. Sixty-four of these 99 reads were further assembled into five contigs with lengths ranging from 123 nt to 225 nt. Two maximum likelihood phylogenetic trees containing these contigs together with partial

sequences of the *cytochrome oxidase I* gene of 38 *Dorylus* sp. specimens (representing all six recognized *Dorylus* subgenera) indicated that four contigs (contigs #1, #3, #4 and #5) clustered with *Dorylus sjoestedti*, *Dorylus wilverthi* and *Dorylus rubellus* whereas the last contig (contig #2) clustered with *Dorylus opacus*, *Dorylus kohli* and *Dorylus helvolus* (Supplementary Figure 1). While contig #2 comprised seven reads obtained from two samples of colony P6, contigs #1, #3, #4 and #5 were composed of 57 reads obtained from 23 samples from 15 colonies (Table 1 and Supplementary Table 1).

In addition, 266 reads that shared identity with the *cytochrome oxidase I* gene were recovered from four Nanopore sequencing runs. These Nanopore reads shared high identity with *Dorylus sjoestedti* and *Dorylus wilverthi* suggesting that ant samples from the FM1 (not assigned by Illumina reads), FM2 and P17 colonies were also related to these two army ant species. Finally, no *cytochrome oxidase I* reads were identified from eleven ant colonies that, therefore, remained taxonomically unassigned (Supplementary Table 1).

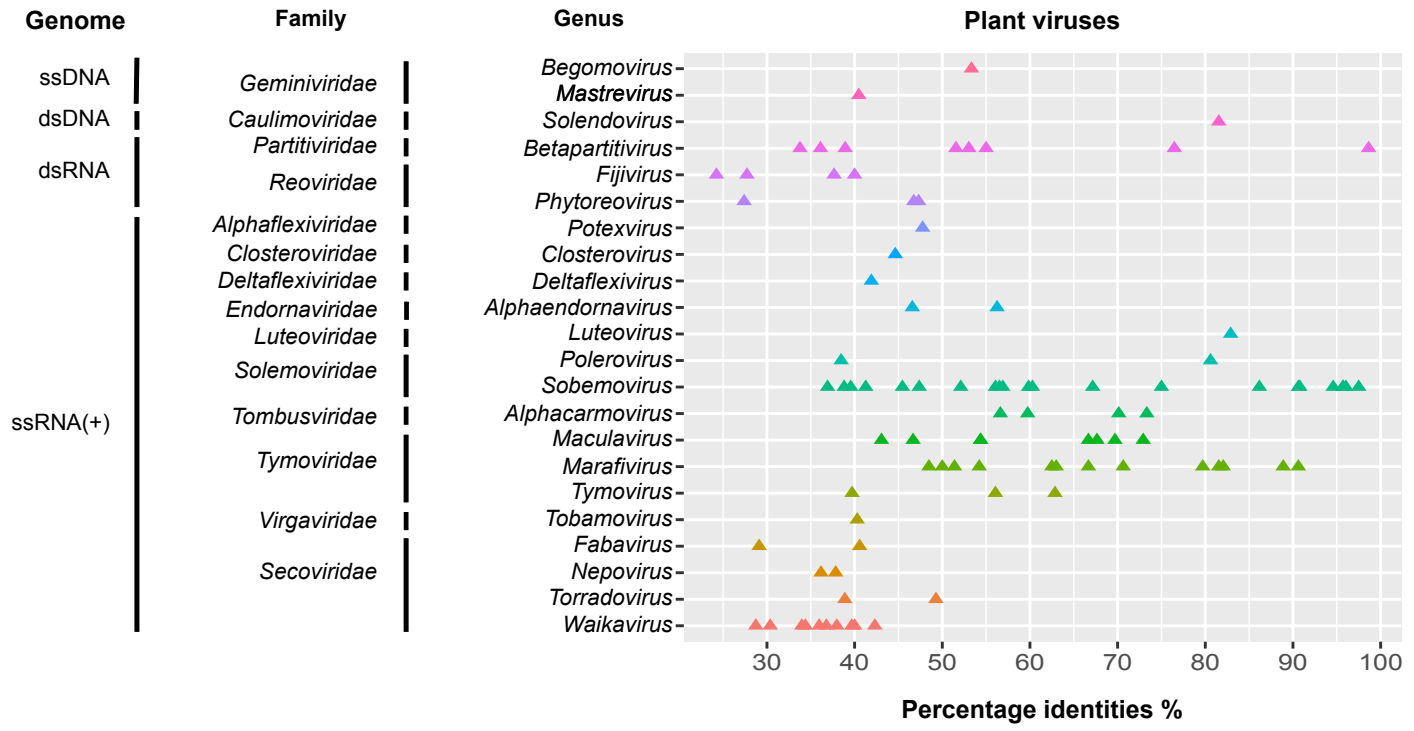
Dorylus sjoestedti, *Dorylus wilverthi* and *Dorylus rubellus* are army ant species that are considered "swarm foragers" (most commonly known as driver ants) on the forest floor and in the lower vegetation [37]. In contrast, *Dorylus opacus* and *Dorylus kohli* which hunt small invertebrates and worms in the leaf-litter are more closely related to subterranean species of *Dorylus* sp., than to the surface swarm foraging ants [37]. Overall, the mitochondrial *cytochrome oxidase I* gene analysis confirmed that the army ants collected in our study were mostly of the genus *Dorylus* and that they were therefore suitable for testing our overarching hypothesis that army ants hunting live invertebrate and vertebrate prey in the deep forest ingest and accumulate numerous plant and animal viruses present in an area around their temporary nests.

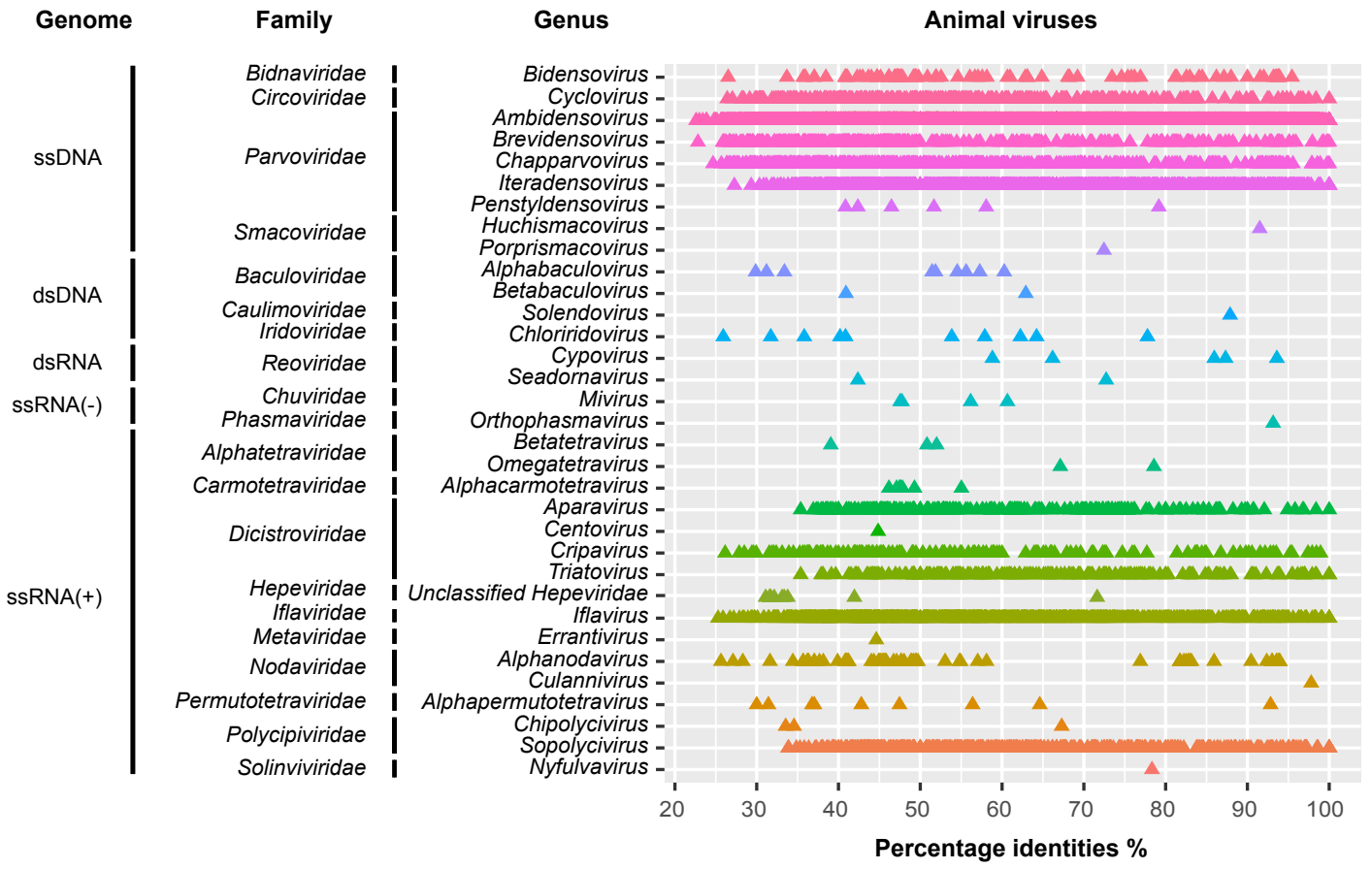
Genetic and morphological factors influence the army ant virome

The three sequencing libraries each contained an average of 50 million reads following removal of short (<15 nt) sequences and individual sequence regions with low quality scores. Overall, 443,645 contigs that were ≥200 nt in length were assembled and 46,377 of these contigs (10.5%) exhibited sequence similarity to viruses (BLASTx e-values < 0.001). However, only 22,406 of these 46,377 contigs (48.3%) exhibited sequence similarity to viral genera recognized or in the process of being recognized by the International Committee on Taxonomy of Viruses (BLASTx e-values < 0.001). These apparently virus-derived contigs yielded detectable homology to viruses of 157 different viral genera in 56 viral families (Supplementary Table 2); overall, a higher diversity than the 29 virus families identified from queens from 49 colonies of North American red fire ant *Solenopsis invicta* [45]. It is noteworthy that the most frequently amplified viral sequences belonged to families containing viruses that are known to infect invertebrates, including *Circoviridae*, *Dicistroviridae*, *Iflaviridae*, *Polycipiviridae* and *Parvoviridae* (Figure 1, Supplementary Table 2). However, plant-, bacteria- and vertebrate-associated viruses were also detected (Figure 1, Supplementary Table 2).

The taxonomic assignations of virus contigs ≥200 nt in length at the genus level were scored for individual ants belonging to 20 *Dorylus* colonies, including 67 individual workers and 78 individual soldiers. No significant difference between the number of virus genera identified from individual ants of the 20 *Dorylus* colonies were found (p-value = 0.1067), suggesting that our roadside sampling survey in the Ogooué-Ivindo region of Gabon was reasonably homogeneous. Remarkably, ants belonging to the *Dorylus opacus/kohli* clade yielded contigs with homology to viruses in significantly fewer genera (median = 4.5, SD = 2.32) than ants assigned to the *Dorylus sjoestedti/wilverthi* clade (median = 9, SD = 4.02; p-value = 0.0057).

Figure 1


Genome
Family
Genus
Plant viruses
ssDNA
dsDNA
dsRNA
ssRNA(+)
Geminiviridae
Caulimoviridae
Partitiviridae
Reoviridae
Alphaflexiviridae
Closteroviridae
Deltaflexiviridae
Endornaviridae
Luteoviridae
Solemoviridae
Tombusviridae
Tymoviridae
Virgaviridae
Secoviridae
Begomovirus
Mastrevirus
Solendovirus
Betapartitivirus
Fijivirus
Phytoreovirus
Potexvirus
Closterovirus
Deltaflexivirus
Alphaendornavirus
Luteovirus
Polerovirus
Sobemovirus
Alphacarmovirus
Maculavirus
Marafivirus
Tymovirus
Tobamovirus
Fabavirus
Nepovirus
Torradovirus
Waikavirus
30
40
50
60
70
80
90
100
Percentage identities %


Genome
Family
Genus
Animal viruses
ssDNA
dsDNA
dsRNA
ssRNA(-)
ssRNA(+)
Bidnaviridae
Circoviridae
Parvoviridae
Smacoviridae
Baculoviridae
Caulimoviridae
Iridoviridae
Reoviridae
Chuviridae
Phasmaviridae
Alphatetraviridae
Carmotetraviridae
Dicistroviridae
Hepeviridae
Iflaviridae
Metaviridae
Nodaviridae
Permutotetraviridae
Polycipiviridae
Solinviviridae
Bidensovirus
Cyclovirus
Ambidensovirus
Brevidensovirus
Chapparvovirus
Iteradensovirus
Penstyldensovirus
Huchismacovirus
Porprismacovirus
Alphabaculovirus
Betabaculovirus
Solendovirus
Chloriridovirus
Cypovirus
Seadornavirus
Mivirus
Orthophasmavirus
Betatetravirus
Omegatetravirus
Alphacarmotetravirus
Aparavirus
Centovirus
Cripavirus
Triatovirus
Unclassified Hepeviridae
Iflavirus
Errantivirus
Alphanodavirus
Culannivirus
Alphapermutotetravirus
Chipolycivirus
Sopolycivirus
Nyfulvavirus
20
30
40
50
60
70
80
90
100
Percentage identities %

While ant species in the *Dorylus opacus/kohli* clade hunt in leaf-litter and forage on forest floors and on vegetation, the driver ant species from the *Dorylus sjoestedti/wilverthi* clade are swarm raiders [25] that prey on invertebrates and occasionally on vertebrates [37]. These driver ants therefore potentially hunt a wider diversity of prey that may result in the intake of a larger diversity of viruses. This result needs to be further confirmed, however, since only eight ants from the *Dorylus opacus/kohli* clade could be analyzed in this study (Table 1).

Finally, we found that contigs obtained from worker ants that were homologous to known viruses represented significantly fewer virus genera (median = 7, SD = 3.52) than those obtained from soldier ants (median = 10, SD = 4.27; p-value = 0.00007). Whereas worker ants specialize in hunting and collecting food, soldier ants have powerful mandibles and specialize in colony defence. In addition, the workers have smaller body sizes than soldier ants. At this stage, we have no clear explanation as to why the diversity of viruses was higher in the soldier ants than in the worker ants. We can hypothesise that i) soldiers simply eat more than workers, ii) soldiers ingest a greater diversity of organisms than workers because some of the animals that are attacked during defence are not necessarily food for the colony, and/or iii) have an immune system that is more permissive of viral infections. Of relevance is also the fact that host body size frequently shows a positive correlation with parasite species richness [46,47], large-bodied soldier ants may provide greater cellular capacity for viral storage/replication than small-body worker ants. Whatever the explanation, this observation suggests that driver-ant soldiers may be ideal candidates for the surveillance of viruses circulating in tropical forest ecosystems.

A broad diversity of bacterial, plant, invertebrate and vertebrate viruses is accumulated by army ants

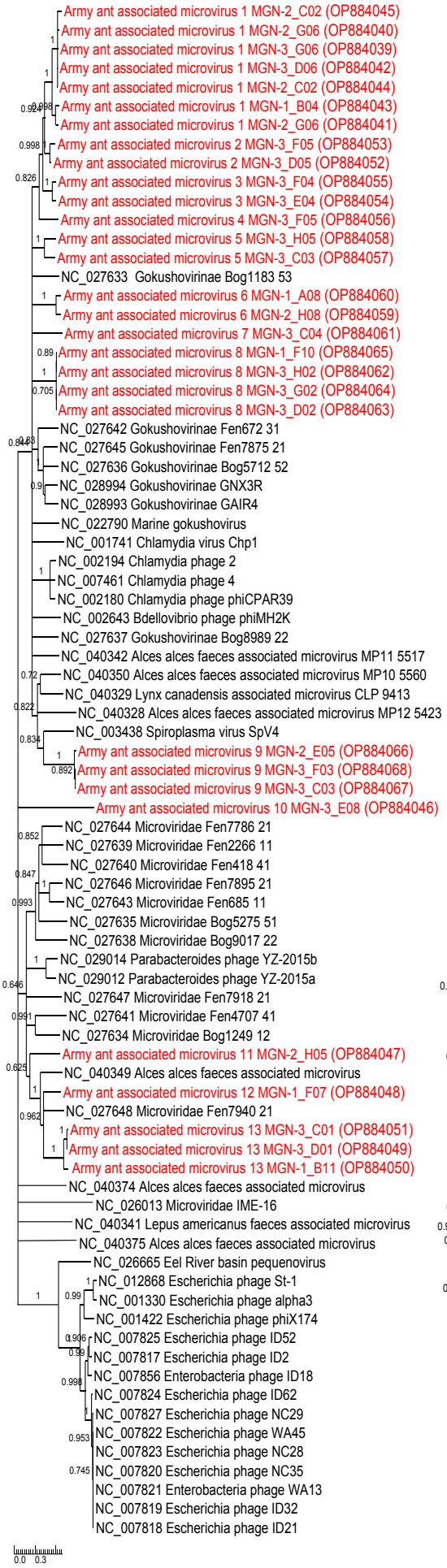
- **Bacteriophages**

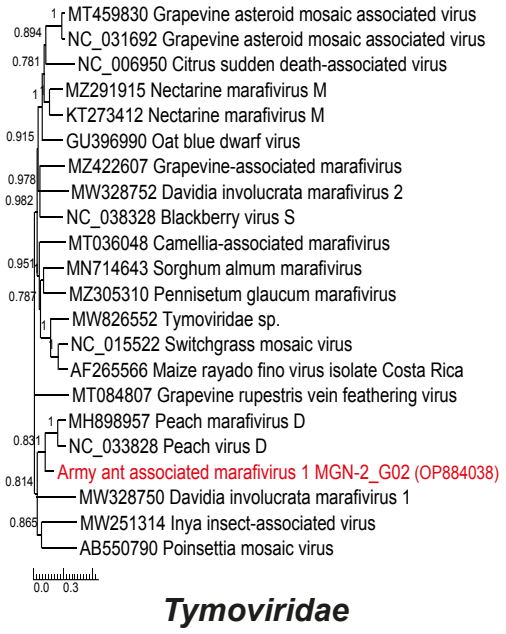
Overall, 583 contigs ≥200 nt in length were assigned to five bacteriophage families (*Herelleviridae*, *Microviridae*, *Namaviridae*, *Picobirnaviridae* and *Tectiviridae*) and three "unclassified Caudoviricetes" subfamilies (recently abolished *Myoviridae*, *Podoviridae* and *Siphoviridae* families [8]) with *Microviridae* accounting for 83.7% (488) of these: a result suggesting that the VANA approach may not efficiently detect bacteriophages with a head and tail structure (Supplementary Table 2). Phylogenetic analyses of the major capsid protein sequences of 30 contigs assigned to the *Microviridae* revealed that these could be subdivided into three groups (Figure 2): (1) one comprising 21 contigs branching with an unclassified *Gokushovirinae* isolate (isolate Bog1183 53; accession number: YP_009160331) recovered from a sphagnum-peat soil [48]; (2) one comprising four contigs clustering with Spiroplasma virus SpV4; Acc. Nbr: NP_598320); and (3) one comprising five contigs grouping with two unclassified microviruses (Alces alces faeces associated microvirus MP21 4718 isolated from moose feces [Acc. Nbr: YP_009551424] and Fen7940 21 [Acc. Nbr: YP_009160412] isolated from a sphagnum-peat soil).

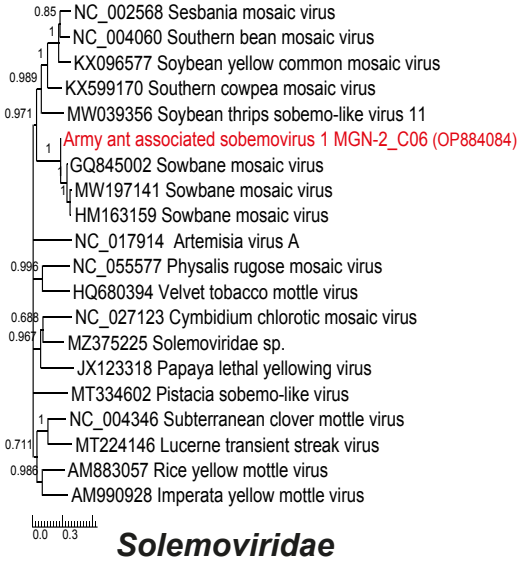
Additionally, 33 contigs were assigned to the *Picobirnaviridae* family (Supplementary Table 2) that is comprised of viruses putatively infecting prokaryotes [49]. Our phylogenetic analyses showed that three army ant contigs (for which we obtained RdRp sequences >130 aa) were unambiguously nested within the picobirnavirus clade (Supplementary Figure 2).

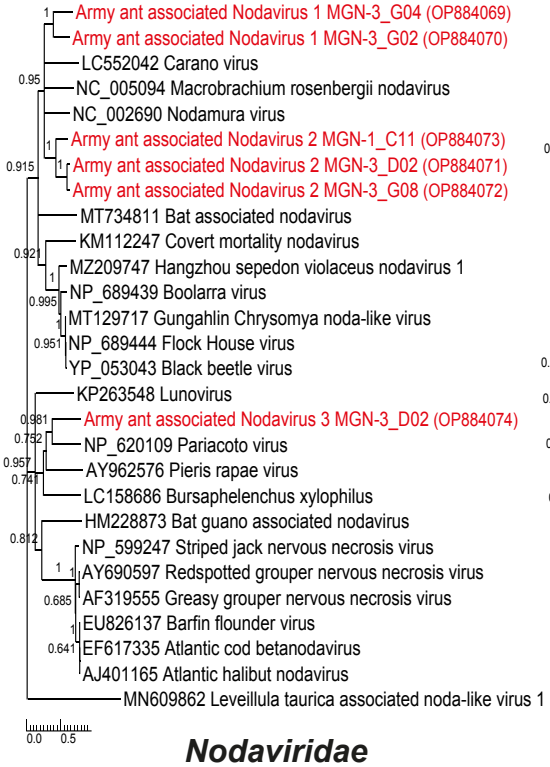
A total of 152 contigs were assigned to the crucivirus group (Supplementary Table 2): a growing group of viruses that appear to have originated through recombination between circular Rep-encoding ssDNA (CRESS-DNA) viruses and RNA tombusviruses [40]. Phylogenetic analyses of the predicted coat protein sequences of three crucivirius-like contigs (for which we

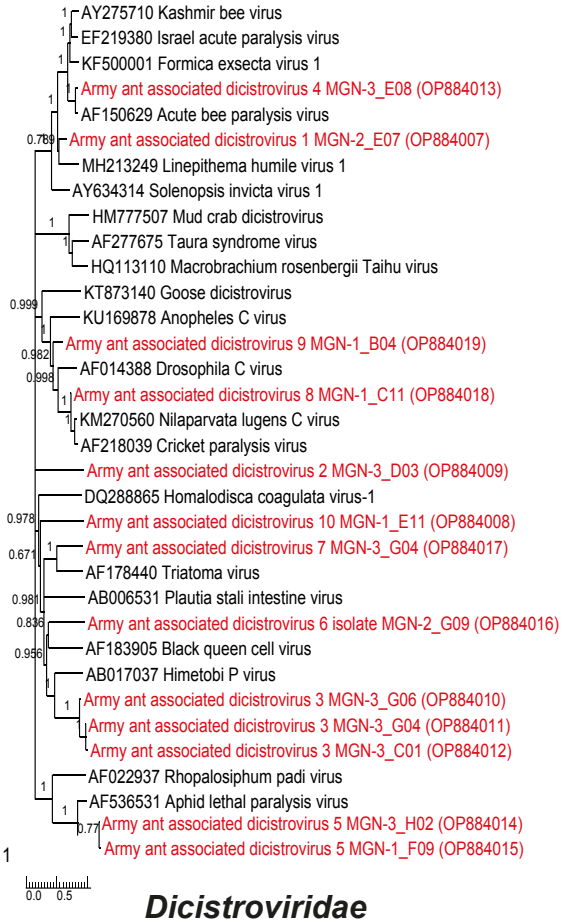
Figure 2


Army ant associated microvirus 1 MGN-2_C02 (OP884045)
Army ant associated microvirus 1 MGN-2_G06 (OP884040)
Army ant associated microvirus 1 MGN-3_G06 (OP884039)
Army ant associated microvirus 1 MGN-3_D06 (OP884042)
Army ant associated microvirus 1 MGN-2_C02 (OP884044)
Army ant associated microvirus 1 MGN-1_B04 (OP884043)
Army ant associated microvirus 1 MGN-2_G06 (OP884041)
Army ant associated microvirus 2 MGN-3_F05 (OP884053)
Army ant associated microvirus 2 MGN-3_D05 (OP884052)
Army ant associated microvirus 3 MGN-3_F04 (OP884055)
Army ant associated microvirus 3 MGN-3_E04 (OP884054)
Army ant associated microvirus 4 MGN-3_F05 (OP884056)
Army ant associated microvirus 5 MGN-3_H05 (OP884058)
Army ant associated microvirus 5 MGN-3_C03 (OP884057)
NC_027633 Gokushovirinae Bog1183 53
Army ant associated microvirus 6 MGN-1_A08 (OP884060)
Army ant associated microvirus 6 MGN-2_H08 (OP884059)
Army ant associated microvirus 7 MGN-3_C04 (OP884061)
Army ant associated microvirus 8 MGN-1_F10 (OP884065)
Army ant associated microvirus 8 MGN-3_H02 (OP884062)
Army ant associated microvirus 8 MGN-3_G02 (OP884064)
Army ant associated microvirus 8 MGN-3_D02 (OP884063)
NC_027642 Gokushovirinae Fen672 31
NC_027645 Gokushovirinae Fen7875 21
NC_027636 Gokushovirinae Bog5712 52
NC_028994 Gokushovirinae GNX3R
NC_028993 Gokushovirinae GAIR4
NC_022790 Marine gokushovirus
NC_001741 Chlamydia virus Chp1
NC_002194 Chlamydia phage 2
NC_007461 Chlamydia phage 4
NC_002180 Chlamydia phage phiCPAR39
NC_002643 Bdellovibrio phage phiMH2K
NC_027637 Gokushovirinae Bog8989 22
NC_040342 Alces alces faeces associated microvirus MP11 5517
NC_040350 Alces alces faeces associated microvirus MP10 5560
NC_040329 Lynx canadensis associated microvirus CLP 9413
NC_040328 Alces alces faeces associated microvirus MP12 5423
NC_003438 Spiroplasma virus SpV4
Army ant associated microvirus 9 MGN-2_E05 (OP884066)
Army ant associated microvirus 9 MGN-3_F03 (OP884068)
Army ant associated microvirus 9 MGN-3_C03 (OP884067)
Army ant associated microvirus 10 MGN-3_E08 (OP884046)
NC_027644 Microviridae Fen7786 21
NC_027639 Microviridae Fen2266 11
NC_027640 Microviridae Fen418 41
NC_027646 Microviridae Fen7895 21
NC_027643 Microviridae Fen685 11
NC_027635 Microviridae Bog5275 51
NC_027638 Microviridae Bog9017 22
NC_029014 Parabacteroides phage YZ-2015b
NC_029012 Parabacteroides phage YZ-2015a
NC_027647 Microviridae Fen7918 21
NC_027641 Microviridae Fen4707 41
NC_027634 Microviridae Bog1249 12
Army ant associated microvirus 11 MGN-2_H05 (OP884047)
NC_040349 Alces alces faeces associated microvirus
Army ant associated microvirus 12 MGN-1_F07 (OP884048)
NC_027648 Microviridae Fen7940 21
Army ant associated microvirus 13 MGN-3_C01 (OP884051)
Army ant associated microvirus 13 MGN-3_D01 (OP884049)
Army ant associated microvirus 13 MGN-1_B11 (OP884050)
NC_040374 Alces alces faeces associated microvirus
NC_026013 Microviridae IME-16
NC_040341 Lepus americanus faeces associated microvirus
NC_040375 Alces alces faeces associated microvirus
NC_026665 Eel River basin pequenovirus
NC_012868 Escherichia phage St-1
NC_001330 Escherichia phage alpha3
NC_001422 Escherichia phage phiX174
NC_007825 Escherichia phage ID52
NC_007817 Escherichia phage ID2
NC_007856 Enterobacteria phage ID18
NC_007824 Escherichia phage ID62
NC_007827 Escherichia phage NC29
NC_007822 Escherichia phage WA45
NC_007823 Escherichia phage NC28
NC_007820 Escherichia phage NC35
NC_007821 Enterobacteria phage WA13
NC_007819 Escherichia phage ID32
NC_007818 Escherichia phage ID21
0.0 0.3
Microviridae


MT459830 Grapevine asteroid mosaic associated virus
NC_031692 Grapevine asteroid mosaic associated virus
NC_006950 Citrus sudden death-associated virus
MZ291915 Nectarine marafivirus M
KT273412 Nectarine marafivirus M
GU396990 Oat blue dwarf virus
MZ422607 Grapevine-associated marafivirus
MW328752 Davidia involucrata marafivirus 2
NC_038328 Blackberry virus S
MT036048 Camellia-associated marafivirus
MN714643 Sorghum almum marafivirus
MZ305310 Pennisetum glaucum marafivirus
MW826552 Tymoviridae sp.
NC_015522 Switchgrass mosaic virus
AF265566 Maize rayado fino virus isolate Costa Rica
MT084807 Grapevine rupestris vein feathering virus
MH898957 Peach marafivirus D
NC_033828 Peach virus D
Army ant associated marafivirus 1 MGN-2_G02 (OP884038)
MW328750 Davidia involucrata marafivirus 1
MW251314 Inya insect-associated virus
AB550790 Poinsettia mosaic virus
0.0 0.3
Tymoviridae


NC_002568 Sesbania mosaic virus
NC_004060 Southern bean mosaic virus
KX096577 Soybean yellow common mosaic virus
KX599170 Southern cowpea mosaic virus
MW039356 Soybean thrips sobemo-like virus 11
Army ant associated sobemovirus 1 MGN-2_C06 (OP884084)
GQ845002 Sowbane mosaic virus
MW197141 Sowbane mosaic virus
HM163159 Sowbane mosaic virus
NC_017914 Artemisia virus A
NC_055577 Physalis rugose mosaic virus
HQ680394 Velvet tobacco mottle virus
NC_027123 Cymbidium chlorotic mosaic virus
MZ375225 Solemoviridae sp.
JX123318 Papaya lethal yellowing virus
MT334602 Pistacia sobemo-like virus
NC_004346 Subterranean clover mottle virus
MT224146 Lucerne transient streak virus
AM883057 Rice yellow mottle virus
AM990928 Imperata yellow mottle virus
0.0 0.3
Solemoviridae


Army ant associated Nodavirus 1 MGN-3_G04 (OP884069)
Army ant associated Nodavirus 1 MGN-3_G02 (OP884070)
LC552042 Carano virus
NC_005094 Macrobrachium rosenbergii nodavirus
NC_002690 Nodamura virus
Army ant associated Nodavirus 2 MGN-1_C11 (OP884073)
Army ant associated Nodavirus 2 MGN-3_D02 (OP884071)
Army ant associated Nodavirus 2 MGN-3_G08 (OP884072)
MT734811 Bat associated nodavirus
KM112247 Covert mortality nodavirus
MZ209747 Hangzhou sepedon violaceus nodavirus 1
NP_689439 Boolarra virus
MT129717 Gungahlin Chrysomya noda-like virus
NP_689444 Flock House virus
YP_053043 Black beetle virus
KP263548 Lunovirus
Army ant associated Nodavirus 3 MGN-3_D02 (OP884074)
NP_620109 Pariacoto virus
AY962576 Pieris rapae virus
LC158686 Bursaphelenchus xylophilus
HM228873 Bat guano associated nodavirus
NP_599247 Striped jack nervous necrosis virus
AY690597 Redspotted grouper nervous necrosis virus
AF319555 Greasy grouper nervous necrosis virus
EU826137 Barfin flounder virus
EF617335 Atlantic cod betanodavirus
AJ401165 Atlantic halibut nodavirus
MN609862 Leveillula taurica associated noda-like virus 1
0.0 0.5
Nodaviridae


AY275710 Kashmir bee virus
EF219380 Israel acute paralysis virus
KF500001 Formica exsecta virus 1
Army ant associated dicistrovirus 4 MGN-3_E08 (OP884013)
AF150629 Acute bee paralysis virus
Army ant associated dicistrovirus 1 MGN-2_E07 (OP884007)
MH213249 Linepithema humile virus 1
AY634314 Solenopsis invicta virus 1
HM777507 Mud crab dicistrovirus
AF277675 Taura syndrome virus
HQ113110 Macrobrachium rosenbergii Taihu virus
KT873140 Goose dicistrovirus
KU169878 Anopheles C virus
Army ant associated dicistrovirus 9 MGN-1_B04 (OP884019)
AF014388 Drosophila C virus
Army ant associated dicistrovirus 8 MGN-1_C11 (OP884018)
KM270560 Nilaparvata lugens C virus
AF218039 Cricket paralysis virus
Army ant associated dicistrovirus 2 MGN-3_D03 (OP884009)
DQ288865 Homalodisca coagulata virus-1
Army ant associated dicistrovirus 10 MGN-1_E11 (OP884008)
Army ant associated dicistrovirus 7 MGN-3_G04 (OP884017)
AF178440 Triatoma virus
AB006531 Plautia stali intestine virus
Army ant associated dicistrovirus 6 isolate MGN-2_G09 (OP884016)
AF183905 Black queen cell virus
AB017037 Himetobi P virus
Army ant associated dicistrovirus 3 MGN-3_G06 (OP884010)
Army ant associated dicistrovirus 3 MGN-3_G04 (OP884011)
Army ant associated dicistrovirus 3 MGN-3_C01 (OP884012)
AF022937 Rhopalosiphum padi virus
AF536531 Aphid lethal paralysis virus
Army ant associated dicistrovirus 5 MGN-3_H02 (OP884014)
Army ant associated dicistrovirus 5 MGN-1_F09 (OP884015)
0.0 0.5
Dicistroviridae

obtained CP sequences >440 aa) and representative sequences of cruciviruses showed that the ant-derived sequences clustered with two unclassified cruciviruses (Crucivirus 250 [Acc. Nbr: MT263579] and Crucivirus 268 [Acc. Nbr: MT263584]) isolated from water and sediments in New Zealand (Supplementary Figure 2).

- **Plant viruses**

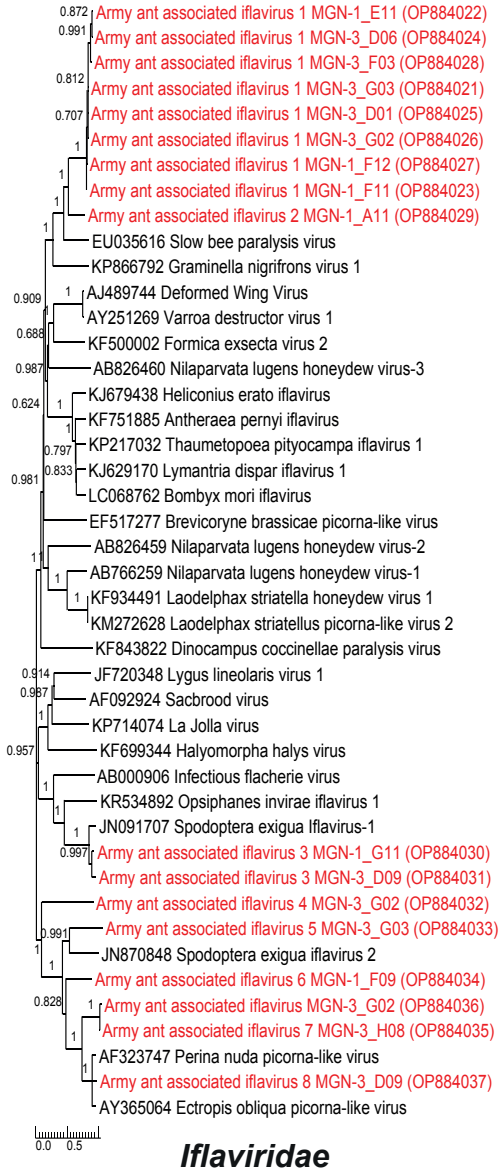
Overall, 101 contigs ≥200 nt in length were assigned to 22 genera in 14 plant virus families (Figure 1, Supplementary Table 2). The size of these contigs ranged from 203 nt to 573 nt (mean = 298 nt), suggesting that they originated from degraded plant virus nucleic acids. This degradation could likely be a consequence of digestion processes along the trophic chain that ended in their eventual presence within the sampled army ants. Plant virus contigs were recovered from 30/145 individual ants (20.7%), including 22/78 soldiers (28.2%) and 8/67 workers (11.9%), suggesting that, as with other viruses, plant viruses tended to accumulate more in soldiers. Interestingly, possible translations of several contigs shared high amino acid identity (86-98%) with two well studied plant viruses: Peach virus D (PeVD; *Tymoviridae* family; *Marafivirus* genus) and Sowbane mosaic virus (SoMV; *Solemoviridae* family; *Sobemovirus* genus). While PeVD-like and SoMV-like contigs (≥ 200 nt) were respectively recovered from four and five ant colonies, smaller (100 - 200 nt) PeVD-like and SoMV-like contigs were, respectively identified from 11 and 9 colonies. These two plant viruses thus appear highly prevalent in the hunting areas of the sampled army ants. Mapping of Illumina and Nanopore reads and contigs assigned to PeVD and SoMV against reference genomes enabled the assembly of partial genome scaffolds respectively covering 39% (2576/6612 nt) and 79.4% (3163/3983 nt) of the PeVD and SoMV full length genomes. Phylogenetic analyses of the SoMV-like RdRp sequences and the PeVD-like polyprotein sequences confirmed that these army ant-associated viruses respectively clustered with known SoMV and PeVD isolates

(Figure 2). Besides detecting these two viruses, contigs with detectable homology with plant virus families containing economically relevant crop pathogens (e.g., *Geminiviridae*, *Solemoviridae*, and *Alphaflexiviridae* Figure 1) were also identified in this study and in a recent study focusing on North American red fire ant *Solenopsis invicta* [45]. This confirms that a broad diversity of plant viruses can potentially be detected within the viromes of top-end predators like army ants that feed on a wide range of herbivorous insects such as whiteflies, aphids, leafhoppers and thrips.

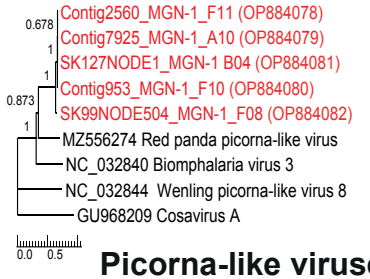
- **Invertebrate-infecting viruses**

Besides recovering a diverse array of vertebrate-, plant- and prokaryote-associated virus sequences, we also identified a large number of highly diverse contigs with detectable homology to known invertebrate-infecting viruses. These included single-strand positive-sense RNA viruses of the *Nodaviridae* (82 contigs), *Dicistroviridae* (643 contigs) and *Iflaviridae* (1067 contigs) families (Supplementary Table 2 and Figures 2 and 3), branching with previously known viral sequences associated with or infecting invertebrates. We also found 1273 contigs sharing detectable homology with the order *Picornavirales* (Supplementary Table 2), including unclassified Picorna-like viruses (793 contigs) and viruses assigned to the *Sopolycivirus* and *Chipolycivirus* genera of the *Polycipiviridae* family (480 contigs; Figure 3), two genera of non-segmented, linear, positive-sense RNA viruses that have previously been found associated with ants (family *Formicidae*) [50]. Finally, we recovered 103 contigs that were assigned to the *Bidnaviridae* family: a new virus family that is composed of bipartite ssDNA viruses infecting invertebrates. Interestingly, it has been suggested that bidnaviruses may have arisen by integration of an ancestral parvovirus genome into a large virus-derived DNA transposon from the polinton family [51]. We here propose that the bidnavirus-like contigs

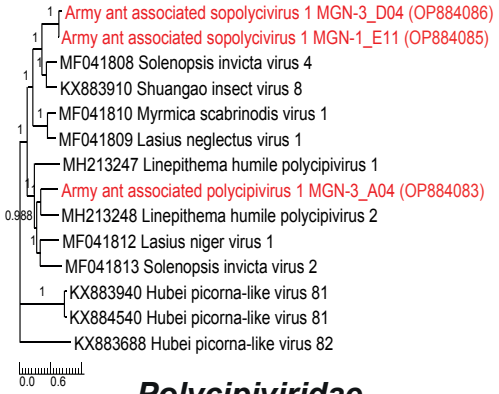
Figure 3



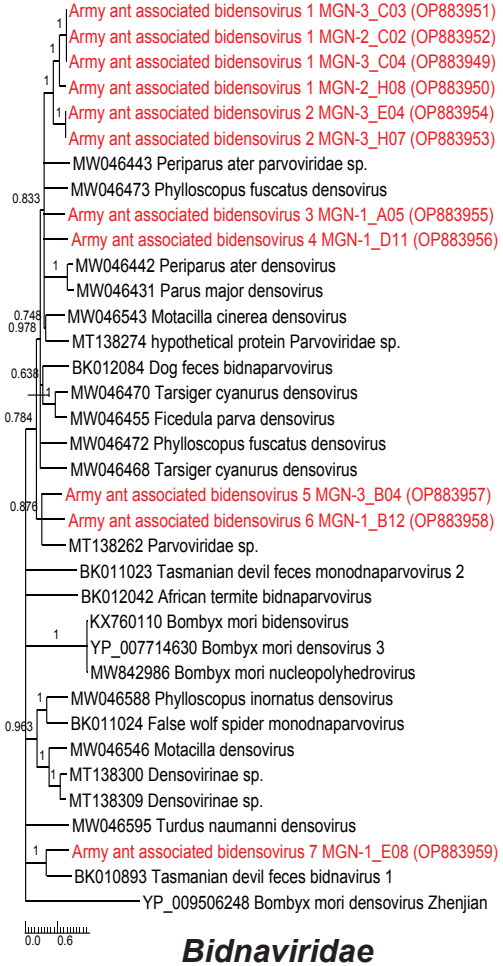
Iflaviridae



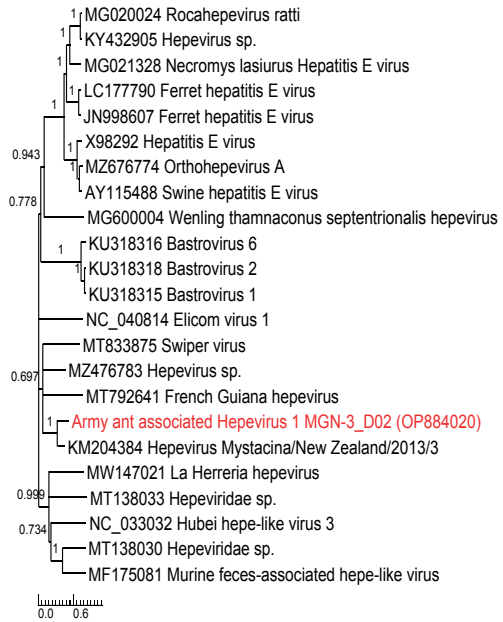
Picorna-like viruses



Polycipiviridae



Bidnaviridae



Hepeviridae

recovered from army ants in our study are likely derived from diverse bipartite bidnaviruses, because these contigs are scattered all through the bidnavirus phylogenetic tree (Figure 3).

- **Vertebrate-infecting viruses**

Because *Dorylus* species have been reported to occasionally prey on vertebrates and scavenge on vertebrate carcasses, we logically investigated the presence of vertebrate viruses. Besides contigs assigned to cyclovirus and parvovirus taxonomic groups that are known to infect vertebrates (see below for a detailed analysis of these two taxa), we found 12 contigs that shared high nucleotide identity with the *Hepeviridae* family, including 4 contigs assigned to the *Orthohepevirus* genus (Supplementary Table 2). The family *Hepeviridae* includes five genera whose members infect salmonid fish, mammals and birds [52]. Specifically, a 2017 nt long contig shared 72.73% identity (e-value = 1.10^{-70}) with Hepevirus Mystacina/New Zealand/2013/3 (Acc. Nbr: KM204384), a hepevirus that was discovered associated with New Zealand lesser short-tailed bats (*Mystacina tuberculata*) [53]. This result was further confirmed by phylogenetic analyses (Figure 3), and supports the hypothesis that the viromes of army ants can include viruses derived from the vertebrates upon which they prey and /or scavenge.

Foraging army ants are reservoirs of a large diversity of parvoviruses and cycloviruses

We focused on two viral families that have been extensively studied in recent years in the context of metagenomics studies: *Parvoviridae* and *Circoviridae*. Both are ssDNA virus families with member species sharing many biological, epidemiological and ecological characteristics. During the last decade, parvoviruses (in particular chapparvoviruses) and circoviruses (in particular cycloviruses) have been characterized from various fluid (e.g. blood) and excretion samples (e.g. feces) from vertebrates [54-57]. Some of these studies have further

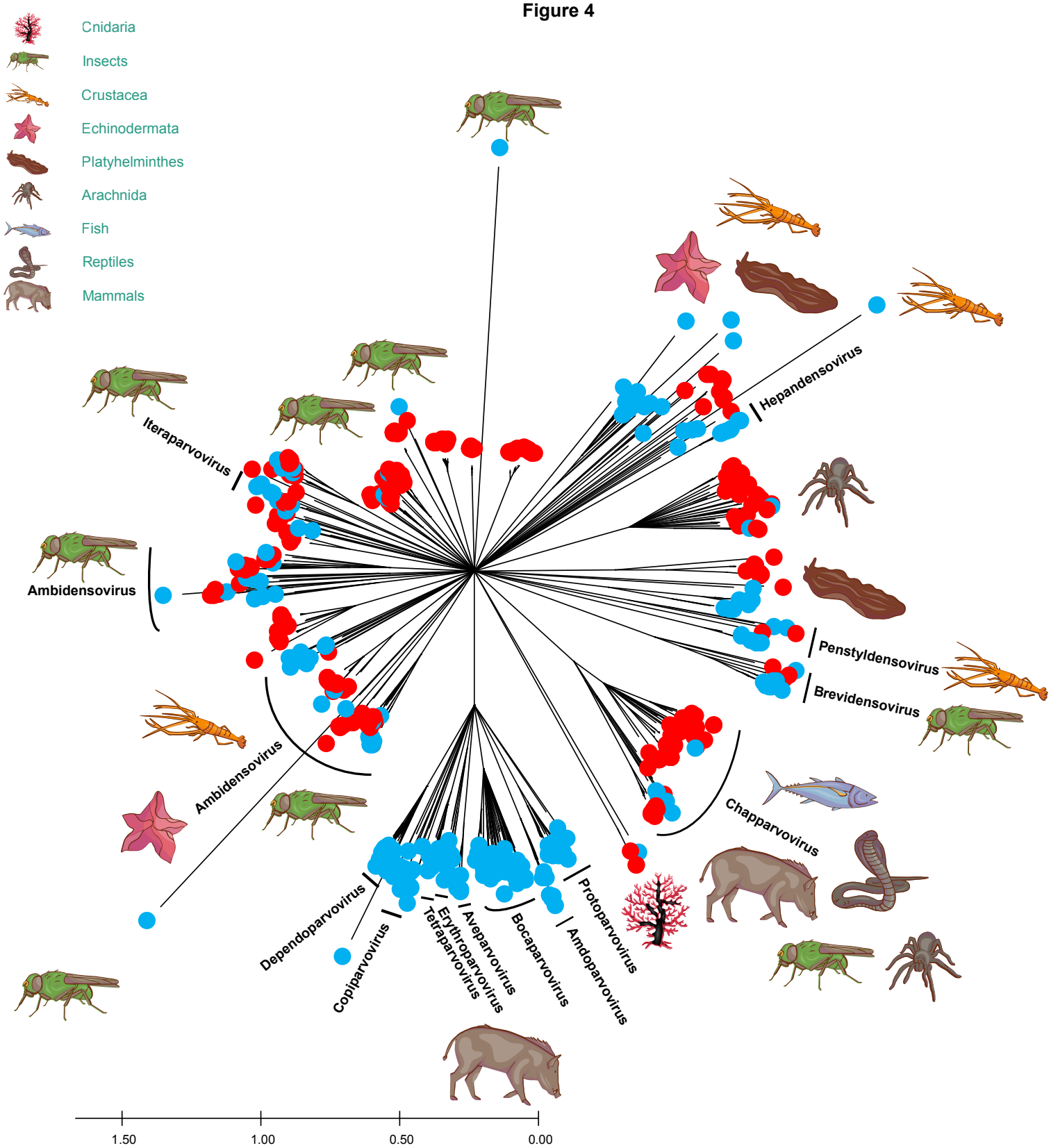
demonstrated that some chapparvoviruses and cycloviruses can actually infect vertebrate hosts [55-57]. However, due to the wide distribution of both viral families also in invertebrates, their natural host-ranges and the pathological characteristics of these viruses remain largely uncertain. For this reason, we were particularly interested in determining the phylogenetic relationships between parvovirus- and cyclovirus sequences found in army ants and those of close relatives that have been sampled previously from vertebrates and invertebrates.

- **Parvoviruses**

The family *Parvoviridae* is composed of animal-infecting viruses which can collectively infect almost all major vertebrate clades and both proto- and deuterostome invertebrates [58]. Parvovirus-related sequences (PRS) were by far the most abundant virus sequences amplified from the army ant samples and accounted for 17,419 contigs (77.8% of all virus contigs).

Four hundred and three army ant-associated SF3 helicase domain sequences (a subdomain of the parvovirus NS1 protein) were scattered around the parvovirus SF3 phylogenetic tree and clustered with SF3 sequences from viruses in almost all parvovirus genera with the notable exception of those in genera of the *Parvovirinae* subfamily (Figure 4). This suggests that a range of animals infected by a broad diversity of parvoviruses was likely preyed or scavenged upon by army ant colonies. These animals were likely predominantly invertebrates including arthropods, molluscs, annelids, nematodes, and cnidarians (Figure 4). This phylogenetic analysis also revealed parvoviruses that might infect army ants. Specifically, some clades contained only army ant-derived SF3 sequences (see for instance the clade located between the two ambidensovirus groups, Figure 4) and it is plausible that the parvoviruses from which these sequences were derived may have been directly infecting the army ants.

Figure 4


Cnidaria
Insects
Crustacea
Echinodermata
Platyhelminthes
Arachnida
Fish
Reptiles
Mammals
Iteraparvovirus
Hepandensovirus
Ambidensovirus
Penstyldensovirus
Brevidensovirus
Ambidensovirus
Chapparvovirus
Dependoparvovirus
Copiparvovirus
Tetraparvovirus
Erythroparvovirus
Aveparvovirus
Bocaparvovirus
Amdoparvovirus
Protoparvovirus
1.50
1.00
0.50
0.00

We further focused on the genetic relationships of the NS1 proteins of chapparvovirus isolates because, while these parvoviruses have been primarily identified by metagenomic studies of animal feces, they have also been both isolated from the tissues of vertebrates (including reptiles, mammals, and birds [54,59]) and found as endogenous parvoviral elements (EPVs) within invertebrate genomes [60]. Recently, this parvovirus group was split into two distinct sub-groups corresponding to two newly established genera: *Ichthamaparvovirus* and *Chaphamaparvovirus* [58]. While ichthamaparvoviruses are known to infect fishes and potentially also invertebrates, chaphamaparvoviruses have only so far been found in vertebrates [60].

Among the parvovirus-like contigs from army ants were several lineages of highly diverse chapparvoviruses (Figure 5) that likely reflect the wide range of prey and carrion upon which these ants feed. These apparently chapparvovirus-derived contigs clustered with invertebrate EPVs (*Mesobuthus martensii*, *Catajapyx aquilonaris* and *Nephila pilipes*) and sequences recovered from bird swabs and a human plasma sample (Figure 5). This suggests that the diversity of chapparvoviruses that infect army ants potentially resulted from the ingestion of infected invertebrates and that chapparvoviruses isolated from animal feces or human plasma may have been derived directly from either parasitic or ingested invertebrates

- **Cycloviruses**

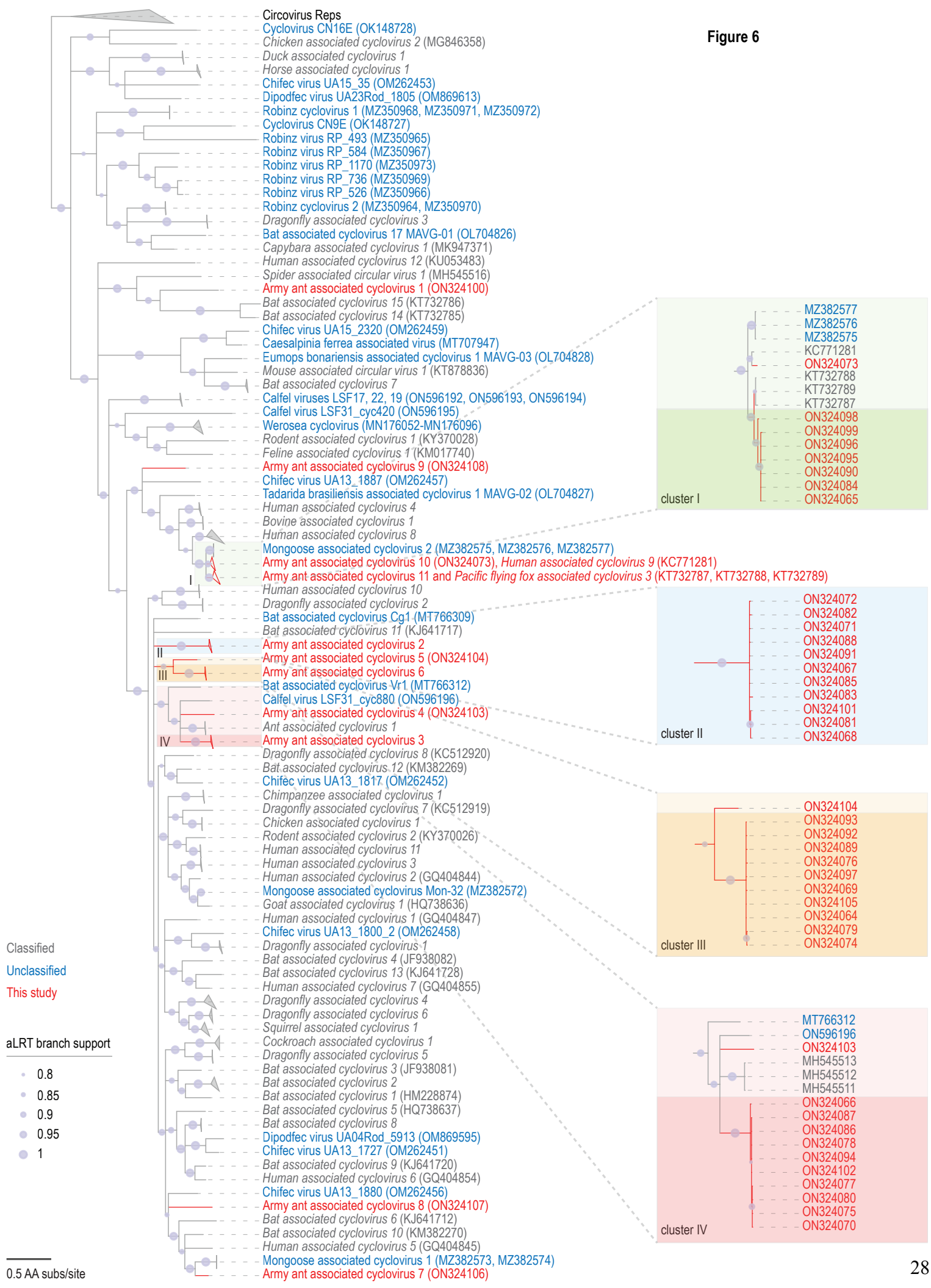
Four hundred and seventy-two contigs ≥ 200 nt in length with detectable homology to sequences of viruses in the family *Circoviridae* (that comprises two genera: *Circovirus* and *Cyclovirus*) clustered in 22 genetic groups (data not shown). Here, we were able to obtain potentially complete genomes of 45 isolates representing the 22 genetic groups from 17 army ant colonies. These genomes, with lengths ranging from 1,723 to 2,024 nt, contained two ORFs predicted to encode a replication-associated protein (Rep) and a capsid (CP) protein. A

Figure 5

KY312550 Parvoviridae sp.
MT732119 Chapparvovirus sp.
MN175613 Psittacara leucophthalmus chapparvovirus
MT138217 MAG: Parvoviridae sp.
MT138295 MAG: Parvoviridae sp.
KF925531 Turkey parvovirus
MT138322 MAG: Parvoviridae sp.
MK513528 Tasmanian devil associated chapparvovirus 1
MN162688 Ichthyic parvovirus
MK448316 Cachavirus
KX272741 Rat parvovirus 2
NC_032097 Desmodus rotundus parvovirus
MF175078 Murine chapparvovirus
MK513529 Tasmanian devil associated chapparvovirus 2
KU563733 Porcine parvovirus 7
MW344047 Sheepavirus Davis 19N
JJRX01097408 WGS: Latrodectus hesperus
JW661431 Fundulus grandis
MN049932 Syngnathus scovelli chapparvovirus
LVHJ00000000 WGS: Hippocampus comes
OM451161 MAG: Hedgehog chapparvovirus 5
AFFK01023528 Strigamia maritima
Army ant associated chapparvovirus 1 MGN-3_D05 (OP883961)
Army ant associated chapparvovirus 1 MGN-3_F05 (OP883960)
JYFJ01046956 WGS: Catajapyx aquilonaris
Army ant associated chapparvovirus 2 MGN-1_E11 (OP883978)
Army ant associated chapparvovirus 2 MGN-1_F11 (OP883977)
Army ant associated chapparvovirus 3 MGN-1_D06 (OP883984)
Army ant associated chapparvovirus 3 MGN-1_D07 (OP883983)
Army ant associated chapparvovirus 3 MGN-2_H08 (OP883982)
Army ant associated chapparvovirus 3 MGN-3_F07 (OP883987)
Army ant associated chapparvovirus 3 MGN-3_G07 (OP883981)
Army ant associated chapparvovirus 3 MGN-3_E07 (OP883985)
Army ant associated chapparvovirus 3 MGN-2_G09 (OP883986)
Army ant associated chapparvovirus 3 MGN-1_C11 (OP883988)
AYEL01046460 WGS: Mesobuthus martensii
Army ant associated chapparvovirus 4 MGN-3_G03 (OP883989)
Army ant associated chapparvovirus 4 MGN-2_G05 (OP883992)
Army ant associated chapparvovirus 4 MGN-2_C05 (OP883991)
Army ant associated chapparvovirus 4 MGN-2_A05 (OP883990)
Army ant associated chapparvovirus 5 MGN-2_C03 (OP883994)
Army ant associated chapparvovirus 5 MGN-1_F09 (OP883995)
Army ant associated chapparvovirus 5 MGN-1_E09 (OP883993)
Army ant associated chapparvovirus 5 MGN-2_A09 (OP883996)
Army ant associated chapparvovirus 6 MGN-2_F10 (OP883997)
Army ant associated chapparvovirus 6 MGN-3_E01 (OP883998)
MN939556 Chapparvovirus sp.
Army ant associated chapparvovirus 7 MGN-3_D06 (OP884000)
Army ant associated chapparvovirus 7 MGN-3_D06 (OP883999)
Army ant associated chapparvovirus 7 MGN-1_A10 (OP884001)
Army ant associated chapparvovirus 8 MGN-3_G06 (OP884002)
Army ant associated chapparvovirus 9 MGN-3_F05 (OP884003)
GFU19424 WGS: Nephila pilipes
Army ant associated chapparvovirus 10 MGN-1_G12 (OP883964)
Army ant associated chapparvovirus 10 MGN-3_F04 (OP883963)
Army ant associated chapparvovirus 10 MGN-3_G06 (OP883962)
Army ant associated chapparvovirus 11 MGN-1_D11 (OP883965)
MT138310 MAG: Parvoviridae sp.
Army ant associated chapparvovirus 12 MGN-1_F09 (OP883966)
Army ant associated chapparvovirus 13 MGN-1_E08 (OP883967)
MW046441 MAG: Periparus ater
Army ant associated chapparvovirus 14 MGN-1_B12 (OP883968)
Army ant associated chapparvovirus 15 MGN-1_A11 (OP883969)
Army ant associated chapparvovirus 16 MGN-3_A04 (OP883971)
Army ant associated chapparvovirus 16 MGN-3_D06 (OP883972)
Army ant associated chapparvovirus 16 MGN-2_F06 (OP883970)
Army ant associated chapparvovirus 16 MGN-3_C05 (OP883973)
Army ant associated chapparvovirus 17 MGN-1_F09 (OP883974)
Army ant associated chapparvovirus 18 MGN-2_E05 (OP883975)
Army ant associated chapparvovirus 19 MGN-1_E05 (OP883976)
Army ant associated chapparvovirus 20 MGN-3_D05 (OP883979)
Army ant associated chapparvovirus 21 MGN-3_H04 (OP883980)
MG846442 Chicken chapparvovirus 2
KM254174 Chicken chapparvovirus HK
MK513533 Tasmanian devil associated chapparvovirus 6
GU371276 Fenneropenaeus chinensis hepandensovirus
Genus Chaphamaparvovirus
Genus Ichthamaparvovirus
Genus Chaphamaparvovirus
1.

conserved nonanucleotide motif (TAGTATTAC) that is typical of the members of the *Circoviridae* family was identified on the *cp*-encoding strand of the 45 genomic sequences. Similarly, all genomes contained two intergenic regions, one located between the 5′ ends of the *rep* and *cp* ORFs, and another between the 3′ ends of the *cp* and *rep* ORFs: again genomic features common in the *Circoviridae* family. Finally, ten out of the 45 *Circoviridae* genome sequences have an apparent intron in the *rep* coding region that is possibly spliced to yield a functional Rep. A phylogenetic analysis showed that the 45 translated Rep sequences are all nested within the phylogenetic tree that included Rep sequences of representative isolates of the *Cyclovirus* genus (Figure 6). Thirty-eight out of the 45 army ant-associated cyclovirus Rep sequences clustered in four groups that respectively contained seven isolates (cluster I in Figure 6), eleven isolates (cluster II in Figure 6), ten isolates (cluster III in Figure 6) and ten isolates (cluster VI in Figure 6). Given that viruses within the *Circoviridae* family are classified into species based on genome-wide pairwise identities with an 80% species demarcation threshold [61], the 45 army ant-associated cycloviruses could reasonably be classified into nine new species (Army ant associated cyclovirus 1, 2, 3, 4, 5, 6, 8 and 10) and two existing *Cyclovirus* species. Specifically, the seven isolates of cluster I (Army ant associated cyclovirus 11) shared 90.8-91.8% identity with Pacific flying fox associated cyclovirus 3 (Acc. Nbr. KT732787, KT732788 and KT732789) and Army ant associated cyclovirus 7 (Acc. Nbr ON324106) shared 80.6% with Bat cyclovirus isolate CyVLysokaP4_CMR_2014 (Acc. Nbr MG693174) which was isolated in Cameroon from fecal samples of *Eidolon helvum*, a fruit-eating bat [62]. Interestingly, a Rep sequence recovered from one of the army ant samples (Army ant associated cyclovirus 10, Acc. Nbr. ON324073) shared 262/278 amino acids (94% identity) with a human cyclovirus (isolate VS5700009, Acc. Nbr. YP_008130363) isolated from a patient with an unexplained paraplegia from Malawi [57]. The CP sequence predicted from the same assembled genome-length contig shared 162/185 aa (88% identity) with an unclassified cyclovirus (isolate

Figure 6


Circovirus Reps
Cyclovirus CN16E (OK148728)
Chicken associated cyclovirus 2 (MG846358)
Duck associated cyclovirus 1
Horse associated cyclovirus 1
Chifec virus UA15_35 (OM262453)
Dipodfec virus UA23Rod_1805 (OM869613)
Robinz cyclovirus 1 (MZ350968, MZ350971, MZ350972)
Cyclovirus CN9E (OK148727)
Robinz virus RP_493 (MZ350965)
Robinz virus RP_584 (MZ350967)
Robinz virus RP_1170 (MZ350973)
Robinz virus RP_736 (MZ350969)
Robinz virus RP_526 (MZ350966)
Robinz cyclovirus 2 (MZ350964, MZ350970)
Dragonfly associated cyclovirus 3
Bat associated cyclovirus 17 MAVG-01 (OL704826)
Capybara associated cyclovirus 1 (MK947371)
Human associated cyclovirus 12 (KU053483)
Spider associated circular virus 1 (MH545516)
Army ant associated cyclovirus 1 (ON324100)
Bat associated cyclovirus 15 (KT732786)
Bat associated cyclovirus 14 (KT732785)
Chifec virus UA15_2320 (OM262459)
Caesalpinia ferrea associated virus (MT707947)
Eumops bonariensis associated cyclovirus 1 MAVG-03 (OL704828)
Mouse associated circular virus 1 (KT878836)
Bat associated cyclovirus 7
Calfel viruses LSF17, 22, 19 (ON596192, ON596193, ON596194)
Calfel virus LSF31_cyc420 (ON596195)
Werosea cyclovirus (MN176052-MN176096)
Rodent associated cyclovirus 1 (KY370028)
Feline associated cyclovirus 1 (KM017740)
Army ant associated cyclovirus 9 (ON324108)
Chifec virus UA13_1887 (OM262457)
Tadarida brasiliensis associated cyclovirus 1 MAVG-02 (OL704827)
Human associated cyclovirus 4
Bovine associated cyclovirus 1
Human associated cyclovirus 8
Mongoose associated cyclovirus 2 (MZ382575, MZ382576, MZ382577)
Army ant associated cyclovirus 10 (ON324073), Human associated cyclovirus 9 (KC771281)
Army ant associated cyclovirus 11 and Pacific flying fox associated cyclovirus 3 (KT732787, KT732788, KT732789)
I
Human associated cyclovirus 10
Dragonfly associated cyclovirus 2
Bat associated cyclovirus Cg1 (MT766309)
Bat associated cyclovirus 11 (KJ641717)
Army ant associated cyclovirus 2
II
Army ant associated cyclovirus 5 (ON324104)
III
Army ant associated cyclovirus 6
Bat associated cyclovirus Vr1 (MT766312)
Calfel virus LSF31_cyc880 (ON596196)
Army ant associated cyclovirus 4 (ON324103)
Ant associated cyclovirus 1
IV
Army ant associated cyclovirus 3
Dragonfly associated cyclovirus 8 (KC512920)
Bat associated cyclovirus 12 (KM382269)
Chifec virus UA13_1817 (OM262452)
Chimpanzee associated cyclovirus 1
Dragonfly associated cyclovirus 7 (KC512919)
Chicken associated cyclovirus 1
Rodent associated cyclovirus 2 (KY370026)
Human associated cyclovirus 11
Human associated cyclovirus 3
Human associated cyclovirus 2 (GQ404844)
Mongoose associated cyclovirus Mon-32 (MZ382572)
Goat associated cyclovirus 1 (HQ738636)
Human associated cyclovirus 1 (GQ404847)
Chifec virus UA13_1800_2 (OM262458)
Dragonfly associated cyclovirus 1
Bat associated cyclovirus 4 (JF938082)
Bat associated cyclovirus 13 (KJ641728)
Human associated cyclovirus 7 (GQ404855)
Dragonfly associated cyclovirus 4
Dragonfly associated cyclovirus 6
Squirrel associated cyclovirus 1
Cockroach associated cyclovirus 1
Dragonfly associated cyclovirus 5
Bat associated cyclovirus 3 (JF938081)
Bat associated cyclovirus 2
Bat associated cyclovirus 1 (HM228874)
Bat associated cyclovirus 5 (HQ738637)
Bat associated cyclovirus 8
Dipodfec virus UA04Rod_5913 (OM869595)
Chifec virus UA13_1727 (OM262451)
Bat associated cyclovirus 9 (KJ641720)
Human associated cyclovirus 6 (GQ404854)
Chifec virus UA13_1880 (OM262456)
Army ant associated cyclovirus 8 (ON324107)
Bat associated cyclovirus 6 (KJ641712)
Bat associated cyclovirus 10 (KM382270)
Human associated cyclovirus 5 (GQ404845)
Mongoose associated cyclovirus 1 (MZ382573, MZ382574)
Army ant associated cyclovirus 7 (ON324106)
Classified
Unclassified
This study
aLRT branch support
0.8
0.85
0.9
0.95
1
0.5 AA subs/site
MZ382577
MZ382576
MZ382575
KC771281
ON324073
KT732788
KT732789
KT732787
ON324098
ON324099
ON324096
ON324095
ON324090
ON324084
ON324065
cluster I
ON324072
ON324082
ON324071
ON324088
ON324091
ON324067
ON324085
ON324083
ON324101
ON324081
ON324068
cluster II
ON324104
ON324093
ON324092
ON324089
ON324076
ON324097
ON324069
ON324105
ON324064
ON324079
ON324074
cluster III
MT766312
ON596196
ON324103
MH545513
MH545512
MH545511
ON324066
ON324087
ON324086
ON324078
ON324094
ON324102
ON324077
ON324080
ON324075
ON324070
cluster IV

ZM36a, Acc. Nbr. YP_009104365) detected within the intestinal contents of a shrew in Zambia [63]. Cyclovirus ZM36a was also closely related to cycloviruses initially identified from human patients with central nervous system manifestations [64]. These results stress the need to better monitor and understand the circulation of cycloviruses between invertebrates, humans and rodents, and the potential of army ants as natural samplers: perhaps even of viruses directly related to human diseases.

Conclusion

This study suggests that predators and scavengers such as army ants can be used to sample broad swathes of environmental viromes including viruses infecting plants, invertebrates and vertebrates. Although not completely unbiased (insects will still have feeding preferences that preclude the sampling of all viruses in an ecosystem), using army ants to sample tropical forest viromes will likely yield sequence data from a more diverse array of viruses than if plants or animals with less diverse diets than army ants were sampled. Army ants will be a particularly good tool for sampling invertebrate viruses in such environments given that they carry what appears to be an extraordinary diversity and abundance of sequences related to known invertebrate-infecting viruses. Although how thoroughly army ants sample the complete invertebrate-associated viromes in the areas surrounding their temporary nests is still unclear, it is undisputable that these top-end predators are probably sampling a non-negligible fraction thereof.

Longitudinal metagenomic analyses of army ant-associated viral nucleic acids in agro-ecological interfaces such as tropical forest areas that bound managed farmlands or human settlements could be a highly convenient means of gaining insights into the ecosystem-scale impacts of natural or human-mediated environmental changes on virus population compositions

and structure. Conversely, monitoring of changes in the relative diversity and prevalence of different viral lineages within the invertebrate-associated virome (and even within plant- and vertebrate-associated viromes) over time could provide a sensitive "leading" indicator of changes in tropical forest ecosystem stability that could foreshadow more obvious changes in these ecosystems due to climate change and other human-mediated disturbances [65].

Acknowledgements

This work has been realized with the support of MESO@LR-Platform at the University of Montpellier.

Funding

No specific funding for this project

Conflict of interest disclosure

The authors declare that they comply with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

Author contributions

S.B., E.L. and P.R. designed the experiment; M.F., P.B., T.N.M., L.B.K., L.H.L., L.B., I.M.M, G.D.M, F.R.N, J.S.K, M.Y, S.B and E.L collected the army ants; M.F., B.R., L.C., E.F., S.K. and J.M.C. performed the molecular biology experiments; M.F., B.R., D.F., F.M., M.O., A.V. and P.R. analyzed the data; B.R., A.V. and P.R. produced the figures; D.P.M. and P.R. wrote the first draft, all authors contributed editorially to the manuscript, and author order was determined on the basis of these duties.

Data, scripts, code, and supplementary information availability

The raw sequence read data from the Illumina HiSeq and Flongle Nanopore runs are available from the NCBI Sequence Read Archive, under the Bioproject accession number PRJNA900057 (Biosample accession numbers SAMN31677877, SAMN31677878, SAMN31677879,

SAMN31726898, SAMN31726899, SAMN31726900 and SAMN31726901) and the following URLs:(https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334884;https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334885;https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334886; https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334634;https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334635;https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334636;https://www.ncbi.nlm.nih.gov/sra/?term=SRR22334637). The 138 army ant associated contig sequences used in the phylogenetic analyses were deposited in GenBank under accession numbers OP883949-OP884086. The 45 army ant associated cyclovirus genome sequences were deposited in GenBank under accession numbers ON324064-ON324108. No script was used to obtain or analyze the result. No codes were used in this study.

References

1 Geoghegan, J. L. & Holmes, E. C. Predicting virus emergence amid evolutionary noise. *Open Biology* **7**, 170189, doi:doi:10.1098/rsob.170189 (2017).

2 Zhang, Y.-Z., Chen, Y.-M., Wang, W., Qin, X.-C. & Holmes, E. C. Expanding the RNA Virosphere by Unbiased Metagenomics. *Annual Review of Virology* **6**, 119-139, doi:10.1146/annurev-virology-092818-015851 (2019).

3 Mokili, J. L., Rohwer, F. & Dutilh, B. E. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* **2**, 63-67 (2012).

4 Wasik, B. R. & Turner, P. E. On the biological success of viruses. *Annu Rev Microbiol* **67**, 519-541, doi:10.1146/annurev-micro-090110-102833 (2013).

5 Greninger, A. L. A decade of RNA virus metagenomics is (not) enough. *Virus Research* **244**, 218-229, doi:10.1016/j.virusres.2017.10.014 (2018).

6 Harvey, E. & Holmes, E. C. Diversity and evolution of the animal virome. *Nature Reviews Microbiology*, doi:10.1038/s41579-021-00665-x (2022).

7 Wren, J. D. *et al.* Plant virus biodiversity and ecology. *PLoS Biol* **4**, e80 (2006).

8 Walker, P. J. *et al.* Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). *Arch. Virol.* **167**, 2429-2440, doi:10.1007/s00705-022-05516-5 (2022).

9 Elena, S. F., Fraile, A. & Garcia-Arenal, F. Evolution and emergence of plant viruses. *Advances in virus research* **88**, 161-191, doi:10.1016/B978-0-12-800098-4.00003-9 (2014).

10 French, R. K. & Holmes, E. C. An Ecosystems Perspective on Virus Evolution and Emergence. *Trends Microbiol* **28**, 165-175, doi:10.1016/j.tim.2019.10.010 (2020).

11 Sommers, P., Chatterjee, A., Varsani, A. & Trubl, G. Integrating Viral Metagenomics into an Ecological Framework. *Annu Rev Virol* **8**, 133-158, doi:10.1146/annurev-virology-010421-053015 (2021).

12 Lefeuvre, P. *et al.* Evolution and ecology of plant viruses. *Nature Reviews Microbiology* **17**, 632-644, doi:10.1038/s41579-019-0232-3 (2019).
13 Poker, J. & MacDicken, K. in *Tropical Forestry Handbook* (eds Laslo Pancel & Michael Köhl) 3-45 (Springer Berlin Heidelberg, 2016).
14 Maclot, F. *et al.* Illuminating an ecological blackbox: Using high throughput sequencing to characterize the plant virome across scales. *Frontiers in Microbiology* **11**, doi:10.3389/fmicb.2020.578064 (2020).
15 Brinkmann, A., Nitsche, A. & Kohl, C. Viral Metagenomics on Blood-Feeding Arthropods as a Tool for Human Disease Surveillance. *International journal of molecular sciences* **17**, 1743, doi:10.3390/ijms17101743 (2016).
16 Fauver, J. R. *et al.* Xenosurveillance reflects traditional sampling techniques for the identification of human pathogens: A comparative study in West Africa. *PLoS Negl Trop Dis* **12**, e0006348, doi:10.1371/journal.pntd.0006348 (2018).
17 Grubaugh, N. D. *et al.* Xenosurveillance: a novel mosquito-based approach for examining the human-pathogen landscape. *PLoS Negl Trop Dis* **9**, e0003628, doi:10.1371/journal.pntd.0003628 (2015).
18 Temmam, S., Davoust, B., Berenger, J. M., Raoult, D. & Desnues, C. Viral metagenomics on animals as a tool for the detection of zoonoses prior to human infection? *International journal of molecular sciences* **15**, 10377-10397, doi:10.3390/ijms150610377 (2014).
19 Ng, T. F. F. *et al.* Exploring the Diversity of Plant DNA Viruses and Their Satellites Using Vector-Enabled Metagenomics on Whiteflies. *Plos One* **6**, e19050 (2011).
20 Rosario, K. *et al.* Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). *The Journal of general virology* **93**, 2668-2681, doi:10.1099/vir.0.045948-0 (2012).
21 Dayaram, A. *et al.* Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infection Genetics and Evolution* **39**, 304-316, doi:10.1016/j.meegid.2016.02.011 (2016).
22 Brady, S. G., Fisher, B. L., Schultz, T. R. & Ward, P. S. The rise of army ants and their relatives: diversification of specialized predatory doryline ants. *BMC Evol Biol* **14**, 93, doi:10.1186/1471-2148-14-93 (2014).
23 Kronauer, D. J. C. Recent advances in army ant biology (Hymenoptera: Formicidae). *Myrmecol News* **12**, 51-65 (2009).
24 Schöning, C., Shepard, L., Sen, A., Kinuthia, W. & Ogutu, J. Should I stay or should I go now? Patch use by African army ant colonies in relation to food availability and predation. *Insectes Sociaux* **58**, 335-343, doi:10.1007/s00040-011-0172-z (2011).
25 Chandra, V., Gal, A. & Kronauer, D. J. C. Colony expansions underlie the evolution of army ant mass raiding. *Proceedings of the National Academy of Sciences of the United States of America* **118** (2021).
26 Powell, S. How much do army ants eat? On the prey intake of a neotropical top-predator. *Insectes Sociaux* **58**, 317-324 (2011).
27 Powell, S. How much do army ants eat? On the prey intake of a neotropical top-predator. *Insectes Sociaux* **58**, 317-324, doi:10.1007/s00040-011-0152-3 (2011).
28 Schöning, C., Njagi, W. & Kinuthia, W. Prey spectra of two swarm-raiding army ant species in East Africa. *Journal of Zoology* **274**, 85-93, doi:https://doi.org/10.1111/j.1469-7998.2007.00360.x (2008).
29 Moubset, O. *et al.* Virion-associated nucleic acid-based metagenomics: a decade of advances in molecular characterization of plant viruses. *Phytopathology*, doi:10.1094/phyto-03-22-0096-rvw (2022).

30 François, S., Filloux, D., Fernandez, E., Ogliastro, M. & Roumagnac, P. Viral metagenomics approaches for high-resolution screening of multiplexed arthropod and plant viral communities. *Methods in molecular biology* **1746**, 77-95, doi:10.1007/978-1-4939-7683-6_7 (2018).
31 Wu, S. & Manber, U. in *Usenix Winter 1992 Technical Conference.* 153-162.
32 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17(1)**, 10 (2011).
33 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).
34 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
35 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666-2669, doi:10.1093/bioinformatics/bty149 (2018).
36 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature methods* **12**, 59-60 (2015).
37 Kronauer, D. J. C., Schöning, C., Vilhelmsen, L. B. & Boomsma, J. J. A molecular phylogeny of Dorylus army ants provides evidence for multiple evolutionary transitions in foraging niche. *BMC Evolutionary Biology* **7**, 56, doi:10.1186/1471-2148-7-56 (2007).
38 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
39 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).
40 de la Higuera, I. *et al.* Unveiling Crucivirus Diversity by Mining Metagenomic Data. *mBio* **11**, doi:10.1128/mBio.01410-20 (2020).
41 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).
42 Cotmore, S. F. *et al.* The family Parvoviridae. *Arch Virol* **159**, 1239-1247, doi:10.1007/s00705-013-1914-1 (2014).
43 François, S. *et al.* Discovery of parvovirus-related sequences in an unexpected broad range of animals. *Scientific reports* **6**, 30880, doi:10.1038/srep30880 (2016).
44 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).
45 Brahma, A., Leon, R. G., Hernandez, G. L. & Wurm, Y. Larger, more connected societies of ants have a higher prevalence of viruses. *Molecular Ecology* **31**, 859-865, doi:https://doi.org/10.1111/mec.16284 (2022).
46 Kamiya, T., O'Dwyer, K., Nakagawa, S. & Poulin, R. What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biol Rev Camb Philos Soc* **89**, 123-134, doi:10.1111/brv.12046 (2014).
47 Poulin, R. Phylogeny, Ecology, and the Richness of Parasite Communities in Vertebrates. *Ecological Monographs* **65**, 283-302, doi:10.2307/2937061 (1995).
48 Quaiser, A. *et al.* Diversity and comparative genomics of Microviridae in Sphagnum-dominated peatlands. *Frontiers in Microbiology* **6**, doi:10.3389/fmicb.2015.00375 (2015).
49 Krishnamurthy, S. R. & Wang, D. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology* **516**, 108-114, doi:10.1016/j.virol.2018.01.006 (2018).

50 Olendraite, I. *et al.* ICTV Virus Taxonomy Profile: Polycipiviridae. *J. Gen. Virol.* **100**, 554-555, doi:https://doi.org/10.1099/jgv.0.001241 (2019).
51 Krupovic, M. & Koonin, E. V. Evolution of eukaryotic single-stranded DNA viruses of the Bidnaviridae family from genes of four other groups of widely different viruses. *Scientific reports* **4**, 5347, doi:10.1038/srep05347 (2014).
52 Purdy, M. A. *et al.* ICTV Virus Taxonomy Profile: Hepeviridae. *J. Gen. Virol.* **98**, 2645-2646, doi:https://doi.org/10.1099/jgv.0.000940 (2017).
53 Wang, J. *et al.* Discovery of novel virus sequences in an isolated and threatened bat species, the New Zealand lesser short-tailed bat (Mystacina tuberculata). *The Journal of general virology* **96**, 2442-2452, doi:10.1099/vir.0.000158 (2015).
54 Fahsbender, E. *et al.* Chapparvovirus DNA Found in 4% of Dogs with Diarrhea. *Viruses* **11**, doi:10.3390/v11050398 (2019).
55 Phan, T. G. *et al.* Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. *Virology* **482**, 98-104, doi:10.1016/j.virol.2015.03.011 (2015).
56 Roediger, B. *et al.* An Atypical Parvovirus Drives Chronic Tubulointerstitial Nephropathy and Kidney Fibrosis. *Cell* **175**, 530-543.e524, doi:10.1016/j.cell.2018.08.013 (2018).
57 Smits, S. L. *et al.* Novel cyclovirus in human cerebrospinal fluid, Malawi, 2010-2011. *Emerg Infect Dis* **19**, 1511-1513, doi:10.3201/eid1909.130404 (2013).
58 Pénzes, J. J. *et al.* Reorganizing the family Parvoviridae: a revised taxonomy independent of the canonical approach based on host association. *Arch Virol* **165**, 2133-2146, doi:10.1007/s00705-020-04632-4 (2020).
59 Souza, W. M. *et al.* Chapparvoviruses occur in at least three vertebrate classes and have a broad biogeographic distribution. *The Journal of general virology* **98**, 225-229, doi:10.1099/jgv.0.000671 (2017).
60 Pénzes, J. J., de Souza, W. M., Agbandje-McKenna, M. & Gifford, R. J. An Ancient Lineage of Highly Divergent Parvoviruses Infects both Vertebrate and Invertebrate Hosts. *Viruses* **11**, doi:10.3390/v11060525 (2019).
61 Rosario, K. *et al.* Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Arch Virol* **162**, 1447-1463, doi:10.1007/s00705-017-3247-y (2017).
62 Yinda, C. K. *et al.* Cameroonian fruit bats harbor divergent viruses, including rotavirus H, bastroviruses, and picobirnaviruses using an alternative genetic code. *Virus Evol* **4**, vey008, doi:10.1093/ve/vey008 (2018).
63 Sasaki, M. *et al.* Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses. *The Journal of general virology* **96**, 440-452, doi:10.1099/vir.0.071209-0 (2015).
64 Tan le, V. *et al.* Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *mBio* **4**, e00231-00213, doi:10.1128/mBio.00231-13 (2013).
65 Neo, J. P. S. & Tan, B. H. The use of animals as a surveillance tool for monitoring environmental health hazards, human health hazards and bioterrorism. *Veterinary microbiology* **203**, 40-48, doi:https://doi.org/10.1016/j.vetmic.2017.02.007 (2017).
66 Stover, B. C. & Muller, K. F. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11**, 7, doi:10.1186/1471-2105-11-7 (2010).

1
2 **FIGURE LEGENDS**

3 **Figure 1: Sequence identity distribution analysis of viral sequences from the army ant**
4 **virome.** Each dot represents an assembled sequence contig with the corresponding protein
5 identity (best BLASTx e-values < 0.001) to plant (top) and animal virus (bottom) in the
6 GenBank nonredundant database.

7

8 **Figure 2: Phylogenies of the sequences of the viruses in the families *Microviridae*,**
9 ***Solemoviridae*, *Tymoviridae, Nodaviridae* and *Dicistroviridae*.** Sequences in red refer to army
10 ant-associated contigs. Neighbor joining phylogenetic trees were generated using alignments
11 of major capsid protein sequences (*Microviridae*), RNA-dependent RNA polymerase
12 sequences (*Dicistroviridae*, *Nodaviridae*, and *Solemoviridae*) and polyprotein sequences
13 (*Tymoviridae*). One thousand bootstrap replicates were performed to quantify branch support.
14 The scale bar depicts the number of amino acid substitutions per site.

15

16 **Figure 3: Phylogenies of the sequences of the viruses in the families *Iflaviridae*,**
17 ***Bidnaviridae*, *Hepeviridae* and *Polycipiviridae*, as well as Picorna-like viruses.** Sequences in
18 red refer to army ant-associated contigs. Neighbor joining phylogenetic trees were generated
19 using alignments of polyprotein sequences (*Iflaviridae*, *Hepeviridae* and Picorna-like viruses),
20 major capsid protein sequences (*Bidnaviridae*) and ORF5 sequences (*Polycipiviridae*). One
21 thousand bootstrap replicates were performed to quantify branch support. The scale bar depicts
22 the number of amino acid substitutions per site.

23

24 **Figure 4: Phylogeny of the sequences of the viruses in the family *Parvoviridae*.** The
25 Neighbor joining phylogenetic tree was generated using alignments of SF3 protein sequences
26 with 500 bootstrap replicates to quantify branch support. Sequences in red refer to army ant-

associated contigs. Putative hosts of several parvoviruses are depicted at the extremity of the branches. Genera of the *Parvoviridae* family are also indicated. The scale bar depicts the number of amino acid substitutions per site.

Figure 5: Phylogeny of the chapparvoviruses. The maximum likelihood tree was generated using alignments of NS1 protein sequences and the WAG substitution model assuming an estimated proportion of invariant sites (of 0.075) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data (gamma=1.337). Support for internal branches was assessed using the aLRT test (SH-Like). Sequences in red refer to army ant-associated contigs. Putative hosts of several parvoviruses are depicted to the right of the branches. The two genera of the chapparvovirus group are also indicated. The scale bar depicts the number of amino acid substitutions per site.

Figure 6: Phylogeny of the cycloviruses. The maximum likelihood tree was generated using alignments of Rep protein sequences and the LG+I+G substitution model. Support for internal branches was assessed using the aLRT test (SH-Like). Branches in red, grey and blue refer to army ant-associated, classified and unclassified cyclovirus Rep sequences. The scale bar depicts the number of amino acid substitutions per site. Branches with <0.8 aLRT support have been collapsed with TreeGraph 2 [66]. Zoomed-in views of four regions of the phylogenetic tree are depicted on the right side of the figure to better localize isolates of Cluster I (dark green area), Cluster II (light blue area), Cluster III (dark orange area) and Cluster IV (dark red area).

1 **SUPPLEMENTARY MATERIAL**

2 *Supplementary table legends*

3

4 **Supplementary Table 1**: Additional characteristics of army ant samples. "nd" means not
5 determined.

6 **Supplementary Table 2**: Number of virus-derived contigs yielding detectable homology to
7 viruses using BLASTx (e-values < 0.001). Note that the *Myoviridae*, *Podoviridae* and
8 *Siphoviridae* families were recently abolished

9 **Supplementary Table 3**: list of abutting primers used for amplifying the whole genomes of
10 army ant associated cycloviruses

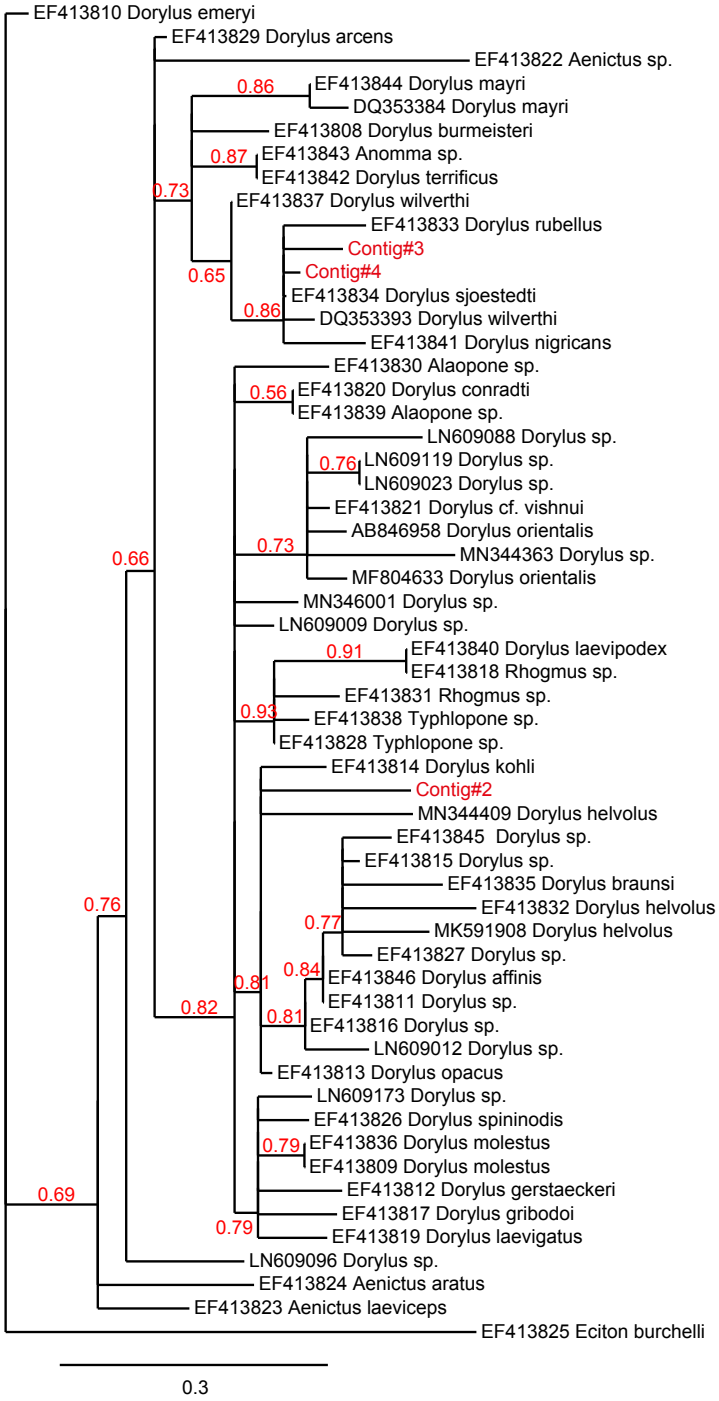
11 *Supplementary figure legends*

12

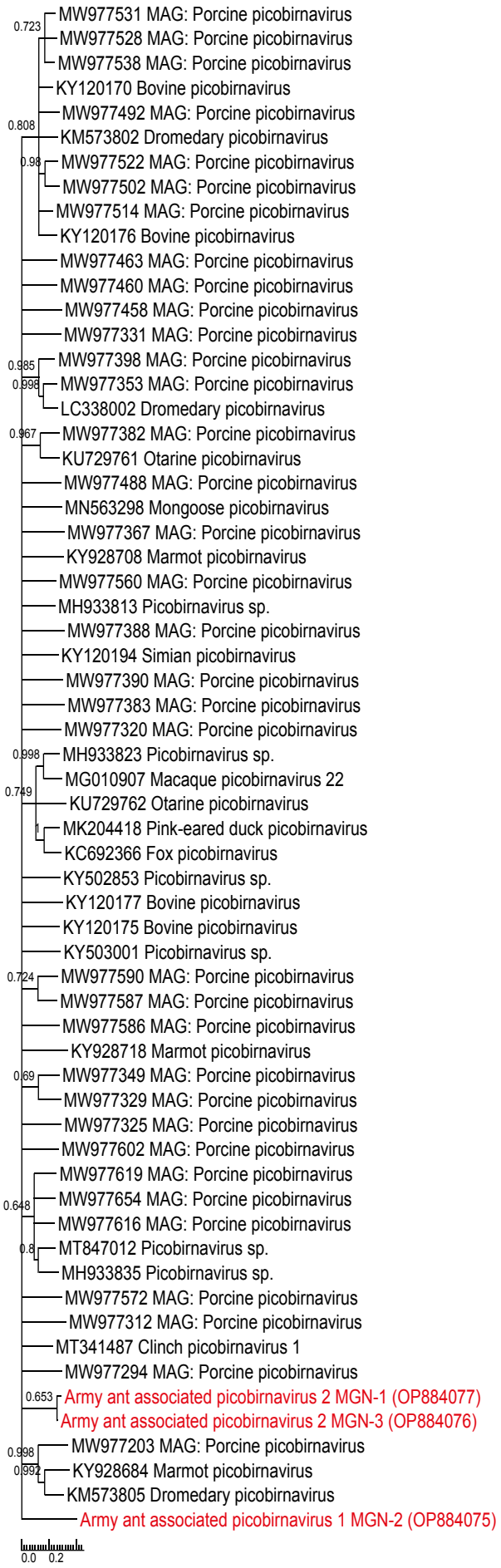
13 **Supplementary Figure 1**: Phylogeny of the Illumina reads assigned to *Dorylus* sp. cytochrome
14 oxidase I gene using BLASTx searches. A phylogenetic tree was constructed using the
15 maximum likelihood method using the HKY85 substitution model assuming an estimated
16 proportion of invariant sites of 0.542 and 4 gamma-distributed rate categories to account for
17 rate heterogeneity across sites. The gamma shape parameter was estimated directly from the
18 data (gamma=1.219). Support for internal branches was assessed using the aLRT test (SH-
19 Like).

20 **Supplementary Figure 2**: Phylogenies of the sequences of the viruses in the *Picobirnaviridae*
21 family, as well as cruciviruses. Sequences in red refer to army ant associated contigs. Neighbor
22 joining phylogenetic trees were generated using alignments of coat protein sequences (CP;
23 cruciviruses) and RNA-dependent RNA polymerase sequences (RdRp; *Picobirnaviridae*). One
24 thousand bootstrap replicates were performed to quantify branch support. The scale bar depicts
25 the number of amino acid substitutions per site.

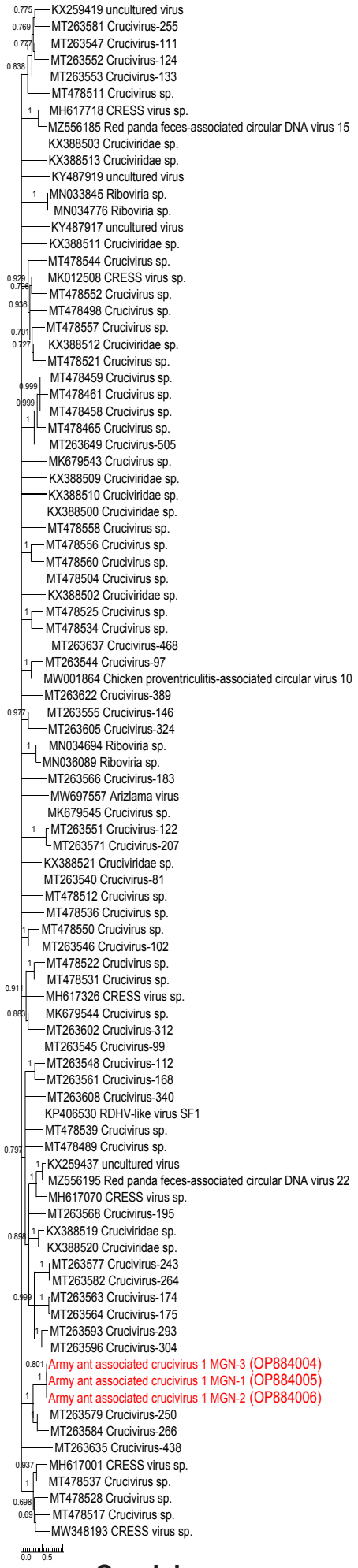
Supplementary Figure 1



Supplementary Figure 2



Picobirnaviridae



Cruciviruses

Supplementary Table 1

Ant colony number	Sample number	Sequencing number	Sample taxonomy	Number of processed ants
FM1	185	MGN-3_B02	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM1	186	MGN-3_C02	nd	1 worker, 1 soldier
FM1	183	MGN-3_G03	nd	1 worker, 1 soldier
FM1	184	MGN-3_H03	nd	1 worker, 1 soldier
FM2	165	MGN-3_A04	nd	1 worker, 1 soldier
FM2	166	MGN-3_B04	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM2	163	MGN-3_G05	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM2	164	MGN-3_H05	nd	1 worker, 1 soldier
FM3	188	MGN-3_D02	nd	1 worker, 1 soldier
FM3	189	MGN-3_E02	nd	1 worker, 1 soldier
FM3	190	MGN-3_F02	nd	1 worker, 1 soldier
FM3	191	MGN-3_G02	nd	1 worker, 1 soldier
FM4	194	MGN-3_B01	nd	1 worker, 1 soldier
FM4	195	MGN-3_C01	nd	1 worker, 1 soldier
FM4	196	MGN-3_D01	nd	1 worker, 1 soldier
FM4	193	MGN-3_H02	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM5	175	MGN-3_A03	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM5	176	MGN-3_B03	nd	1 worker, 1 soldier
FM5	173	MGN-3_G04	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM5	174	MGN-3_H04	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM6	178	MGN-3_C03	nd	1 worker, 1 soldier
FM6	179	MGN-3_D03	nd	1 worker, 1 soldier
FM6	180	MGN-3_E03	nd	1 worker, 1 soldier
FM6	181	MGN-3_F03	nd	1 worker, 1 soldier
FM7	168	MGN-3_C04	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM7	169	MGN-3_D04	nd	1 worker, 1 soldier
FM7	170	MGN-3_E04	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
FM7	171	MGN-3_F04	nd	1 worker, 1 soldier
FM8	198	MGN-3_E01	nd	1 worker, 1 soldier
FM8	199	MGN-3_F01	nd	1 worker, 1 soldier
FM8	200	MGN-3_G01	nd	1 worker, 1 soldier
FM8	201	MGN-3_H01	nd	1 worker, 1 soldier
P10	P10-reste	MGN-1_A05	nd	4 workers, 4 soldiers
P10	P10-1.1	MGN-1_A06	nd	1 worker
P10	P10-1.2	MGN-1_B06	nd	1 soldier
P10	P10-2.1	MGN-1_C06	nd	1 worker
P10	P10-2.2	MGN-1_D06	nd	1 soldier
P10	P10-3.1	MGN-1_E06	nd	1 worker
P10	P10-3.2	MGN-1_F06	nd	1 soldier
P10	P10-4.1	MGN-1_G06	nd	1 worker
P10	P10-4.2	MGN-1_H06	nd	1 soldier
P11	P11-4.2	MGN-1_A04	nd	1 soldier
P11	P11-reste	MGN-1_B04	nd	4 workers, 4 soldiers
P11	P11-1.1	MGN-1_B05	nd	1 worker
P11	P11-1.2	MGN-1_C05	nd	1 soldier
P11	P11-2.1	MGN-1_D05	nd	1 worker
P11	P11-2.2	MGN-1_E05	nd	1 soldier
P11	P11-3.1	MGN-1_F05	nd	1 worker

P11	P11-3.2	MGN-1_G05	nd	1 soldier
P11	P11-4.1	MGN-1_H05	nd	1 worker
P12	P12-1.1	MGN-1_C04	nd	1 worker
P12	P12-1.2	MGN-1_D04	nd	1 soldier
P12	P12-2.1	MGN-1_E04	nd	1 worker
P12	P12-2.2	MGN-1_F04	*D. sjoestedti/wilverthi* clade	1 soldier
P12	P12-3.1	MGN-2_A07	nd	1 worker
P12	P12-3.2	MGN-2_B07	nd	1 soldier
P12	P12-4.1	MGN-2_C07	nd	1 worker
P12	P12-4.2	MGN-2_D07	nd	1 soldier
P12	P12-reste	MGN-2_E07	nd	4 workers, 4 soldiers
P13	P13-2.2	MGN-2_A06	nd	1 soldier
P13	P13-3.1	MGN-2_B06	nd	1 worker
P13	P13-3.2	MGN-2_C06	nd	1 soldier
P13	P13-4.1	MGN-2_E06	nd	1 worker
P13	P13-4.2	MGN-2_F06	nd	1 soldier
P13	P13-1.1	MGN-2_F07	nd	1 worker
P13	P13-reste	MGN-2_G06	nd	4 workers, 4 soldiers
P13	P13-1.2	MGN-2_G07	nd	1 soldier
P13	P13-2.1	MGN-2_H07	nd	1 worker
P14	P14-1.2	MGN-2_A05	nd	1 soldier
P14	P14-2.1	MGN-2_B05	nd	1 worker
P14	P14-2.2	MGN-2_C05	*D. sjoestedti/wilverthi* clade	1 soldier
P14	P14-3.1	MGN-2_D05	nd	1 worker
P14	P14-3.2	MGN-2_E05	nd	1 soldier
P14	P14-4.1	MGN-2_F05	nd	1 worker
P14	P14-4.2	MGN-2_G05	nd	1 soldier
P14	P14-reste	MGN-2_H05	nd	4 workers, 4 soldiers
P14	P14-1.1	MGN-2_H06	nd	1 worker
P15	P15-reste	MGN-2_A03	*D. sjoestedti/wilverthi* clade	4 workers, 4 soldiers
P15	P15-1.1	MGN-2_A04	nd	1 worker
P15	P15-1.2	MGN-2_B04	nd	1 soldier
P15	P15-2.1	MGN-2_C04	nd	1 worker
P15	P15-2.2	MGN-2_D04	nd	1 soldier
P15	P15-3.1	MGN-2_E04	nd	1 worker
P15	P15-3.2	MGN-2_F04	nd	1 soldier
P15	P15-4.1	MGN-2_G04	nd	1 worker
P15	P15-4.2	MGN-2_H04	nd	1 soldier
P16	P16-4.1	MGN-2_A02	nd	1 worker
P16	P16-4.2	MGN-2_B02	nd	1 soldier
P16	P16-1.1	MGN-2_B03	nd	1 worker
P16	P16-reste	MGN-2_C02	*D. sjoestedti/wilverthi* clade	4 workers, 4 soldiers
P16	P16-1.2	MGN-2_C03	nd	1 soldier
P16	P16-2.1	MGN-2_E03	nd	1 worker
P16	P16-2.2	MGN-2_F03	nd	1 soldier
P16	P16-3.1	MGN-2_G03	nd	1 worker
P16	P16-3.2	MGN-2_H03	nd	1 soldier
P17	P17-1.1	MGN-2_D02	nd	1 worker
P17	P17-1.2	MGN-2_E02	nd	1 soldier
P17	P17-2.1	MGN-2_F02	*D. sjoestedti/wilverthi* clade	1 worker
P17	P17-2.2	MGN-2_G02	*D. sjoestedti/wilverthi* clade	1 soldier
P17	P17-3.1	MGN-2_H02	nd	1 worker

P17	P17-3.2	MGN-3_A09	nd	1 soldier
P17	P17-4.1	MGN-3_B09	nd	1 worker
P17	P17-4.2	MGN-3_C09	nd	1 soldier
P17	P17-reste	MGN-3_D09	nd	5 workers, 5 soldiers
P18	P18-3.1	MGN-3_A08	nd	1 worker
P18	P18-3.2	MGN-3_B08	nd	1 soldier
P18	P18-4.1	MGN-3_C08	nd	1 worker
P18	P18-4.2	MGN-3_D08	nd	1 soldier
P18	P18-reste	MGN-3_E08	*D. sjoestedti/wilverthi* clade	4 workers, 4 soldiers
P18	P18-1.1	MGN-3_E09	nd	1 worker
P18	P18-1.2	MGN-3_F09	nd	1 soldier
P18	P18-2.1	MGN-3_G09	nd	1 worker
P18	P18-2.2	MGN-3_H09	nd	1 soldier
P19	P19-2.1	MGN-3_A07	nd	1 soldier
P19	P19-2.2	MGN-3_B07	nd	1 soldier
P19	P19-3.1	MGN-3_C07	nd	1 soldier
P19	P19-3.2	MGN-3_D07	nd	1 soldier
P19	P19-4.1	MGN-3_E07	nd	1 soldier
P19	P19-4.2	MGN-3_F07	nd	1 soldier
P19	P19-reste	MGN-3_G07	nd	1 worker, 7 soldiers
P19	P19-1.1	MGN-3_G08	nd	1 soldier
P19	P19-1.2	MGN-3_H08	nd	1 soldier
P1A	P1A-reste	MGN-1_A11	nd	4 workers, 4 soldiers
P1A	P1A-1.1	MGN-1_A12	nd	1 worker
P1A	P1A-1.2	MGN-1_B12	nd	1 soldier
P1A	P1A-2.1	MGN-1_C12	nd	1 worker
P1A	P1A-2.2	MGN-1_D12	nd	1 soldier
P1A	P1A-3.1	MGN-1_E12	nd	1 worker
P1A	P1A-3.2	MGN-1_F12	nd	1 soldier
P1A	P1A-4.1	MGN-1_G12	nd	1 soldier
P1A	P1A-4.2	MGN-1_H12	nd	1 worker
P21/23	P21/23-1.2	MGN-3_A06	nd	1 soldier
P21/23	P21/23-2.1	MGN-3_B06	nd	1 worker
P21/23	P21/23-2.2	MGN-3_C06	*D. sjoestedti/wilverthi* clade	1 soldier
P21/23	P21/23-reste	MGN-3_D06	*D. sjoestedti/wilverthi* clade	5 workers, 5 soldiers
P21/23	P21/23-1.1	MGN-3_H07	nd	1 worker
P24	P24-larves	MGN-3_A05	nd	several larvae
P24	P24-3.1	MGN-3_B05	nd	1 worker
P24	P24-3.2	MGN-3_C05	nd	1 soldier
P24	P24-4.1	MGN-3_D05	nd	1 soldier
P24	P24-4.2	MGN-3_E05	nd	1 soldier
P24	P24-1.1	MGN-3_E06	nd	1 worker
P24	P24-reste	MGN-3_F05	nd	4 workers, 4 soldiers
P24	P24-1.2	MGN-3_F06	nd	1 soldier
P24	P24-2.1	MGN-3_G06	nd	1 soldier
P24	P24-2.2	MGN-3_H06	nd	1 worker
P2A	P2A-1	MGN-1_B11	nd	1 worker, 1 soldier
P2A	P2A-2	MGN-1_C11	*D. sjoestedti/wilverthi* clade	1 worker, 1 soldier
P2A	P2A-3	MGN-1_D11	nd	1 worker, 1 soldier
P2A	P2A-4	MGN-1_E11	nd	1 worker, 1 soldier
P2A	P2A-reste	MGN-1_F11	nd	13 workers
P3A	P3A-1	MGN-1_A10	nd	1 worker, 2 soldier

P3A	P3A-3.1	MGN-1_B10	nd	2 workers
P3A	P3A-3.2	MGN-1_C10	nd	1 soldier
P3A	P3A-4.1	MGN-1_D10	nd	1 worker
P3A	P3A-4.2	MGN-1_E10	nd	1 soldier
P3A	P3A-reste	MGN-1_F10	nd	4 workers, 4 soldiers
P3A	P3A-2.1	MGN-1_G11	nd	1 worker
P3A	P3A-2.2	MGN-1_H11	*D. sjoestedti/wilverthi* clade	1 soldier
P4A	P4A-2.1	MGN-1_A09	nd	1 worker
P4A	P4A-2.2	MGN-1_B09	nd	1 soldier
P4A	P4A-3.1	MGN-1_C09	nd	1 worker
P4A	P4A-3.2	MGN-1_D09	nd	1 soldier
P4A	P4A-4	MGN-1_E09	nd	1 worker, 1 soldier
P4A	P4A-reste	MGN-1_F09	*D. sjoestedti/wilverthi* clade	4 workers, 4 soldiers
P4A	P4A-1.1	MGN-1_G10	nd	1 worker
P4A	P4A-1.2	MGN-1_H10	*D. sjoestedti/wilverthi* clade	1 soldier
P5	P5-2	MGN-1_A08	nd	1 worker, 1 soldier
P5	P5-3.1	MGN-1_B08	nd	1 worker
P5	P5-3.2	MGN-1_C08	nd	1 soldier
P5	P5-4.1	MGN-1_D08	nd	1 worker
P5	P5-4.2	MGN-1_E08	*D. sjoestedti/wilverthi* clade	1 soldier
P5	P5-reste	MGN-1_F08	nd	4 workers, 4 soldiers
P5	P5-1.1	MGN-1_G09	nd	1 worker
P5	P5-1.2	MGN-1_H09	nd	1 soldier
P6	P6-2.1	MGN-1_A07	nd	1 worker
P6	P6-2.2	MGN-1_B07	nd	1 soldier
P6	P6-3.1	MGN-1_C07	nd	1 worker
P6	P6-3.2	MGN-1_D07	*D. opacus/kohli* clade	1 soldier
P6	P6-4	MGN-1_E07	nd	1 worker, 1 soldier
P6	P6-reste	MGN-1_F07	*D. opacus/kohli* clade	4 workers, 4 soldiers
P6	P6-1.1	MGN-1_G08	nd	1 worker
P6	P6-1.2	MGN-1_H08	nd	1 soldier
P7	P7-1.1	MGN-1_G07	nd	1 worker
P7	P7-1.2	MGN-1_H07	nd	1 soldier
P7	P7-2	MGN-2_A10	nd	1 worker, 1 soldier
P7	P7-3	MGN-2_B10	*D. sjoestedti/wilverthi* clade	1 soldier
P7	P7-4.1	MGN-2_C10	nd	1 worker
P7	P7-4.2	MGN-2_D10	nd	1 worker
P7	P7-4.3	MGN-2_E10	nd	1 worker
P7	P7-reste	MGN-2_F10	nd	2 workers, 3 soldiers
P8A	P8A-3.1	MGN-2_A09	nd	1 worker
P8A	P8A-3.2	MGN-2_B09	nd	1 soldier
P8A	P8A-4.1	MGN-2_C09	nd	1 worker
P8A	P8A-4.2	MGN-2_D09	nd	1 soldier
P8A	P8A-5.1	MGN-2_E09	nd	1 worker
P8A	P8A-5.2	MGN-2_F09	nd	1 soldier
P8A	P8A-reste	MGN-2_G09	nd	4 workers, 5 soldiers
P8A	P8A-2.1	MGN-2_G10	nd	1 worker
P8A	P8A-2.2	MGN-2_H10	nd	1 soldier
P9	P9-1.2	MGN-2_A08	nd	1 soldier
P9	P9-2.1	MGN-2_B08	nd	1 soldier
P9	P9-2.2	MGN-2_C08	nd	1 soldier
P9	P9-3.1	MGN-2_D08	nd	1 worker

P9	P9-3.2	MGN-2_E08	nd	1 soldier
P9	P9-4.1	MGN-2_F08	nd	1 worker
P9	P9-4.2	MGN-2_G08	nd	1 soldier
P9	P9-reste	MGN-2_H08	nd	4 workers, 4 soldiers
P9	P9-1.1	MGN-2_H09	nd	1 worker

Supplementary Table 2

Family	Genus	Nucleic acid	Putative host										Number contigs > 200nt
			Plant	Fungi	Bacteria	Vertebrate	Invertebrate	Archae	Amoeba	Alga	Virus	stramenopiles / alveolates / rhizaria	
Adenoviridae	*Mastadenovirus*	dsDNA				X							4
Alphaflexiviridae	*Potexvirus*	ssRNA (+)	X										1
Alphatetraviridae	*Betatetravirus*	ssRNA (+)					X						3
Alphatetraviridae	*Omegatetravirus*	ssRNA (+)					X						2
Asfarviridae	*Asfivirus*	dsDNA				X							2
Baculoviridae	*Alphabaculovirus*	dsDNA					X						9
Baculoviridae	*Betabaculovirus*	dsDNA					X						2
Bidnaviridae	*Bidensovirus*	ssDNA					X						103
Carmotetraviridae	*Alphacarmotetravirus*	ssRNA (+)					X						8
Caulimoviridae	*Solendovirus*	dsDNA	X				X						1
Chuviridae	*Mivirus*	ssRNA (-)					X						4
Circoviridae	*Cyclovirus*	ssDNA				X	X						303
Circoviridae	*Circovirus*	ssDNA				X							169
Closteroviridae	*Closterovirus*	ssRNA (+)	X										1
Deltaflexiviridae	*Deltaflexivirus*	ssRNA (+)	X	X									1
Dicistroviridae	*Aparavirus*	ssRNA (+)					X						287
Dicistroviridae	*Triatovirus*	ssRNA (+)					X						217
Dicistroviridae	*Cripavirus*	ssRNA (+)					X						138
Dicistroviridae	*Centovirus*	ssRNA (+)					X						1
Endornaviridae	*Alphaendornavirus*	dsRNA	X	X									2
Geminiviridae	*Begomovirus*	ssDNA	X										1
Geminiviridae	*Mastrevirus*	ssDNA	X										1
Genomoviridae	*Gemycircularvirus*	ssDNA		X		X							27
Genomoviridae	*Gemykibivirus*	ssDNA				X							17
Genomoviridae	*Gemygorvirus*	ssDNA				X							4
Genomoviridae	*Gemykolovirus*	ssDNA				X							4
Hepeviridae	*Orthohepevirus*	ssRNA (+)				X							4
Herelleviridae	*Tsarbombavirus*	dsDNA			X								3

Herelleviridae	*Sepunavirus*	dsDNA		X						2
Herelleviridae	*Agatevirus*	dsDNA		X						1
Herelleviridae	*Wphvirus*	dsDNA		X						1
Herpesviridae	*Muromegalovirus*	dsDNA			X					1
Herpesviridae	*Roseolovirus*	dsDNA			X					1
Iflaviridae	*Iflavirus*	ssRNA (+)				X				1067
Iridoviridae	*Chloriridovirus*	dsDNA				X				10
Iridoviridae	*Lymphocystivirus*	dsDNA			X					1
Lavidaviridae	*Sputnikvirus*	dsDNA							X	1
Luteoviridae	*Luteovirus*	ssRNA (+)	X							1
Marseilleviridae	*Marseillevirus*	dsDNA					X			3
Marseilleviridae	*Tunisvirus*	dsDNA					X			1
Mesoniviridae	*Alphamesonivirus*	ssRNA (+)				X				1
Metaviridae	*Errantivirus*	ssRNA (+)				X				1
Microviridae	*Microvirus*	ssDNA		X						462
Microviridae	*Spiromicrovirus*	ssDNA		X						23
Microviridae	*Chlamydiamicrovirus*	ssDNA		X						2
Microviridae	*Bdellomicrovirus*	ssDNA		X						1
Mimiviridae	*Klosneuvirus*	dsDNA					X			11
Mimiviridae	*Mimivirus*	dsDNA					X			8
Mimiviridae	*Tupanvirus*	dsDNA					X			4
Mimiviridae	*Hokovirus*	dsDNA					X			3
Mimiviridae	*Indivirus*	dsDNA					X			3
Mimiviridae	*Moumouvirus*	dsDNA					X			3
Mimiviridae	*Catovirus*	dsDNA					X			1
Myoviridae	*Polybotosvirus*	dsDNA		X						4
Myoviridae	*Ficleduovirus*	dsDNA		X						2
Myoviridae	*Obolenskvirus*	dsDNA		X						2
Myoviridae	*Seunavirus*	dsDNA		X						2
Myoviridae	*Vequintavirus*	dsDNA		X						2
Myoviridae	*Vidavervirus*	dsDNA		X						2
Myoviridae	*Asteriusvirus*	dsDNA		X						1

Myoviridae	*Certrevirus*	dsDNA			X				1
Myoviridae	*Emdodecavirus*	dsDNA			X				1
Myoviridae	*Eneladusvirus*	dsDNA			X				1
Myoviridae	*Gelderlandvirus*	dsDNA			X				1
Myoviridae	*Schizotequatrovirus*	dsDNA			X				1
Myoviridae	*Tegunavirus*	dsDNA			X				1
Narnaviridae	*Narnavirus*	dsDNA			X				1
Nodaviridae	*Alphanodavirus*	ssRNA (+)				X	X		74
Nodaviridae	*Betanodavirus*	ssRNA (+)				X			7
Nodaviridae	*Culannivirus*	ssRNA (+)					X		1
Papillomaviridae	*Gammapapillomavirus*	dsDNA				X			3
Papillomaviridae	*Deltapapillomavirus*	dsDNA				X			2
Papillomaviridae	*Alefpapillomavirus*	dsDNA				X			1
Papillomaviridae	*Alphapapillomavirus*	dsDNA				X			1
Papillomaviridae	*Betapapillomavirus*	dsDNA				X			1
Papillomaviridae	*Dyorhopapillomavirus*	dsDNA				X			1
Partitiviridae	*Betapartitivirus*	dsRNA	X	X					8
Parvoviridae	*Ambidensovirus*	ssDNA					X		15040
Parvoviridae	*Iteradensovirus*	ssDNA					X		1389
Parvoviridae	*Chapparvovirus*	ssDNA				X	X		472
Parvoviridae	*Brevidensovirus*	ssDNA					X		377
Parvoviridae	*Protoparvovirus*	ssDNA				X			76
Parvoviridae	*Bocaparvovirus*	ssDNA				X			21
Parvoviridae	*Aveparvovirus*	ssDNA				X			19
Parvoviridae	*Dependoparvovirus*	ssDNA				X			9
Parvoviridae	*Penstyldensovirus*	ssDNA					X		6
Parvoviridae	*Copiparvovirus*	ssDNA				X			5
Parvoviridae	*Erythroparvovirus*	ssDNA				X			5
Permutotetraviridae	*Alphapermutotetravirus*	ssRNA (+)					X		9
Phasmaviridae	*Orthophasmavirus*	ssRNA (-)					X		1
Phycodnaviridae	*Chlorovirus*	dsDNA						X	28
Phycodnaviridae	*Prasinovirus*	dsDNA						X	14

Phycodnaviridae	*Prymnesiovirus*	dsDNA						X	8
Phycodnaviridae	*Coccolithovirus*	dsDNA						X	1
Phycodnaviridae	*Phaeovirus*	dsDNA						X	1
Picobirnaviridae	*Picobirnavirus*	dsRNA		X	X				33
Picornaviridae	*Livupivirus*	ssRNA (+)			X				3
Picornaviridae	*Enterovirus*	ssRNA (+)			X				2
Picornaviridae	*Megrivirus*	ssRNA (+)			X				2
Picornaviridae	*Hepatovirus*	ssRNA (+)			X				1
Picornaviridae	*Limnipivirus*	ssRNA (+)			X				1
Picornaviridae	*Rabovirus*	ssRNA (+)			X				1
Picornaviridae	*Rosavirus*	ssRNA (+)			X				1
Picornaviridae	*Tremovirus*	ssRNA (+)			X				1
Pithoviridae	*Pithovirus*	dsDNA					X		1
Podoviridae	*Drulisvirus*	dsDNA		X					1
Podoviridae	*Fipvunavirus*	dsDNA		X					1
Podoviridae	*Jwalphavirus*	dsDNA		X					1
Podoviridae	*Kuravirus*	dsDNA		X					1
Polycipiviridae	*Sopolycivirus*	ssRNA (+)				X			477
Polycipiviridae	*Chipolycivirus*	ssRNA (+)				X			3
Polyomaviridae	*Gammapolyomavirus*	dsDNA			X				2
Poxviridae	*Orthopoxvirus*	dsDNA			X				25
Poxviridae	*Parapoxvirus*	dsDNA			X				1
Reoviridae	*Cypovirus*	dsRNA				X			5
Reoviridae	*Fijivirus*	dsRNA	X						4
Reoviridae	*Coltivirus*	dsRNA			X				3
Reoviridae	*Phytoreovirus*	dsRNA	X						3
Reoviridae	*Orbivirus*	dsRNA			X				2
Reoviridae	*Orthoreovirus*	dsRNA			X				2
Reoviridae	*Seadornavirus*	dsRNA			X	X			2
Retroviridae	*Gammaretrovirus*	ssRNA (+)			X				232
Retroviridae	*Betaretrovirus*	ssRNA (+)			X				1
Retroviridae	*Lentivirus*	ssRNA (+)			X				1

Secoviridae	*Waikavirus*	ssRNA (+)	X						10
Secoviridae	*Fabavirus*	ssRNA (+)	X						2
Secoviridae	*Nepovirus*	ssRNA (+)	X						2
Secoviridae	*Torradovirus*	ssRNA (+)	X						2
Siphoviridae	*Skunavirus*	dsDNA			X				7
Siphoviridae	*Cheoctovirus*	dsDNA			X			X	2
Siphoviridae	*Detrevirus*	dsDNA			X				2
Siphoviridae	*Incheonvirus*	dsDNA			X				2
Siphoviridae	*Jerseyvirus*	dsDNA			X				2
Siphoviridae	*Nipunavirus*	dsDNA			X				2
Siphoviridae	*Efquatrovirus*	dsDNA			X				1
Siphoviridae	*Jesfedecavirus*	dsDNA			X				1
Siphoviridae	*Klementvirus*	dsDNA			X				1
Siphoviridae	*Moineauvirus*	dsDNA			X				1
Siphoviridae	*Pahexavirus*	dsDNA			X				1
Siphoviridae	*Pamexvirus*	dsDNA			X				1
Siphoviridae	*Samistivirus*	dsDNA			X				1
Siphoviridae	*Saphexavirus*	dsDNA			X				1
Siphoviridae	*Tunavirus*	dsDNA			X				1
Siphoviridae	*Wilnyevirus*	dsDNA			X				1
Smacoviridae	*Huchismacovirus*	ssDNA				X	X		1
Smacoviridae	*Porprismacovirus*	ssDNA				X	X		1
Solemoviridae	*Sobemovirus*	ssRNA (+)	X						22
Solemoviridae	*Polerovirus*	ssRNA (+)	X						2
Solinviviridae	*Nyfulvavirus*	ssRNA (+)					X		1
Tectiviridae	*Alphatectivirus*	dsDNA			X				1
Tectiviridae	*Gammatectivirus*	dsDNA			X				1
Tombusviridae	*Alphacarmovirus*	ssRNA (+)	X						12
Totiviridae	*Totivirus*	dsRNA		X					5
Tymoviridae	*Marafivirus*	ssRNA (+)	X						13
Tymoviridae	*Maculavirus*	ssRNA (+)	X						8
Tymoviridae	*Tymovirus*	ssRNA (+)	X						3

Virgaviridae	*Tobamovirus*	ssRNA (+)	**X**			1
Not determined	Unclassified cruciviruses	ssDNA			**X**	152
Not determined	Unclassified Picorna-like viruses	ssRNA (+)		**X**		793

Supplementary Table 3

Primer name	Primer sequence
AntvOTU2_F	CGCTGACTCGAATATATACCTAGAGCCCC
AntvOTU2_R	TAGTCTCAGGTTCCTAAGTTGACTCGATGG
AntvOTU3_F	TCCCTGCAGATGTTTTGTTCCAGTAGTAGG
AntvOTU3_R	TTTTGGTAAGTACATGCTTGGACAGTCTGG
AntvOTU4_F	AGAATCTTTTGCGAATTCTTCTGCTTCTGC
AntvOTU4_R	ATCTACTACAAACCAAGAGGACTTTGGTGG
AntvOTU5_F	GAAGATGTCCGAAAGAGTATTATGGAGCCC
AntvOTU5_R	GTTGGGTCGGTGCTGTCATAATTCATTCC
AntvOTU6_F	ACGTTATTTTCAACAGTAACAGAGGCAACC
AntvOTU6_R	TAACAGTTTTTGGGGTGCTAGCTTCTTACC
AntvOTU11_F	AGAGCATTAGAGTCAACATTCTCAATACGCG
AntvOTU11_R	TAATCTCCCGCATTATTGTGCTCTTATCGC
AntvOTU16_F	TTTTGGTAAGTACACGCTTGGATAGTCTGG
AntvOTU16_R	TCCCTGCAGATGTTTTGTTCCAGTAGTAGG
AntvOTU17_F	CCATCAAGAAGCATCTGCATCACTCAATCC
AntvOTU17_R	CGTTAAATCGCATGGGTTTTGTAAGATTGC
AntvOTU19_F	GACAACTGGTACCATTTCACCAACTACACC
AntvOTU19_R	CACTCTGGTGTTGGAGGTGATAATGATTCG
AntvOTU20_F	CCATACAAGGTACAAGTGAAAGGAGGATTCG
AntvOTU20_R	GTAGCGATCCATTATTTTGAGCATTTCGTCG
AntvOTU22_F	ACAAAATCAAACATACTGTCGAAAAGGCGG
AntvOTU22_R	TCGTCTGTCCCGTTTGCCTTTTCAATATGG