Abundant and Diverse RNA Viruses in Insects Revealed by RNA-Seq Analysis: Ecological and Evolutionary Implications

Haoming Wu,[a] Rui Pang,[a] Tong Cheng,[a] Liang Xue,[a] Haiyan Zeng,[a] Tao Lei,[a] Moutong Chen,[a] Shi Wu,[a] Yu Ding,[b] Jumei Zhang,[a] Mang Shi,[c,d] Qingping Wu[a]

[a]State Key Laboratory of Applied Microbiology Southern China, Guangdong Provincial Key Laboratory of Microbiology Culture Collection and Application, Guangdong Open Laboratory of Applied Microbiology, Guangdong Institute of Microbiology, Guangdong Academy of Sciences, Guangzhou, China

[b]Department of Food Science and Technology, Jinan University, Guangzhou, China

[c]School of Medicine, Sun Yat-sen University, Guangzhou, China

[d]School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, New South Wales, Australia

Haoming Wu and Rui Pang contributed equally to this work. Author order was determined by drawing straws.

ABSTRACT Increasing data indicate that insects serve as major reservoirs and vectors of viruses, which account for the continuously increasing ecological burden and infectious disease outbreaks. Uncovering the hidden diversity of viruses in insects will further the understanding of the ecological and evolutionary perspectives in the emergence of insect-associated virus diseases. In this study, we queried transcriptome sequencing (RNA-Seq) data from more than 600 species across 32 insect orders dwelling in different ecological habitats and recovered more than 1,213 RNA viruses that were recapitulated in 40 families, 2 unclassified genera, and many unspecified viral groups. These novel viruses included the well-known insect-associated viruses within *Flaviviridae*, *Picornavirales*, *Bunyavirales*, *Mononegavirales*, *Nidovirales*, *Reoviridae*, and *Negevirus*. More appeared to form novel clusters within previously described taxa or could be resolved as paraphyletic, including the first astrovirus identified in insects, in which many were sufficiently divergent to warrant the establishment of new virus genera or families. Additionally, some viruses were closely related to the recognized plant-, fungus-, and vertebrate-specific species, implying the importance of relationships between insect behavior and virus spread. Comparative genome analyses also revealed high genomic variability with respect to the flexible gene pool and genome architecture of these newly described viruses, including the evidence for genome reshuffling first discovered in *Dicistroviridae*. The data reflecting the genetically and ecologically diverse viral populations in insects greatly expand our understanding of RNA viruses in nature and highlight that the biodiversity of RNA viruses remains largely unexplored.

IMPORTANCE Insects comprise the largest proportion of animals on earth and are frequently implicated in the transmission of vector-borne diseases. However, considerable attention has been paid to the phytophagous and hematophagous insects, with results that provide insufficient and biased information about the viruses in insects. Here, we have delivered compelling evidence for the exceptional abundance and genetic diversity of RNA viruses in a wide range of insects. Novel viruses were found to cover major categories of RNA viruses, and many formed novel clusters divergent from the previously described taxa, dramatically broadening the range of known RNA viruses in insects. These newly characterized RNA viruses exhibited high levels of genomic plasticity in genome size, open reading frame (ORF) number, intergenic structure, and gene rearrangement and segmentation. This work provides comprehensive insight into the origin, spread, and evolution of RNA viruses. Of

course, a large-scale virome project involving more organisms would provide more-detailed information about the virus infections in insects.

Viruses are ubiquitous on earth and are deemed to infect all known cellular organisms across the global ecosystems (1, 2). Historically, virus identification has focused on specific organismal communities, such as humans and clinically and commercially associated animals and plants. However, these represent only a small proportion of the range of viral biodiversity in nature and, to a great extent, the data have led to biased inference in characterizing the evolutionary landscape of viruses. An enormous amount of viruses remains undiscovered using classical viral detection methods due to the inability to cultivate the vast majority of viruses *in vitro*. The application of viral metagenomics has facilitated the identification of unknown viruses in a broad host range (3), from unicellular organisms to mammals, terrestrial or aquatic (4–8). Because of their extreme abundance and adaptability in diverse environments and habitats, animals have been implicated frequently in the evolution and spread of viruses. Recent studies have described abundant and diverse RNA viruses in animals, providing important insights about the host adaptation and distribution of viruses (9–14). Still, the hidden diversity of RNA viruses within the ecologically and geographically diverse animal species remains to be explored.

Insects are arguably the most widespread and diverse group of animals on the planet, occupying different ecological niches and playing a dominant role in the functioning of ecosystems (15). Viral infections have been invoked as a significant threat to many ecologically and commercially important insects, with dramatic declines seen in the species richness and diversity in the agricultural and forest ecosystems (16, 17). For the plant sap- and blood-sucking insects, the food and feeding habits make them hot spots for interspecies transmission and emergence of novel viruses (18–20). Indeed, more than 70% of the known plant viruses depend on insect vectors for their survival and spread (21). In the past decades, the increased incidence of insect-borne epidemics has been documented, including the new emergence of epidemic Zika virus and reemergence of dengue (DENV), chikungunya (CHIKV), yellow fever (YFV), and other infectious diseases (22, 23). These viruses are transmitted between their hosts by the insect vectors, causing no overt pathology during circulation in insect populations (24). Though many more viruses have been persistently identified in insects, little is known about their ecological distribution and evolutionary dynamic in the wild.

Over time, ecological species identification has revealed large and diverse populations of insects, as well as of their parasites (25). As obligate cellular parasites, viruses share intimate relationships with their hosts, raising issues concerning the genetic landscape and macroevolutionary patterns of the fast-evolving RNA viruses in insects. Previous studies focused on those insects with public health and economic importance or on those restricted to specific habitats, such as mosquitoes, honey bees, and flies (13, 26–31). Few studies have been performed on the insects as a whole. As a result of the deluge of omics studies, an unprecedented explosion in biological data (i.e., genomics and transcriptomics) is providing a particularly abundant resource for virus discovery (5, 8, 32). Here, we conducted a systematic study to explore RNA viruses of more than 600 insect species by querying publicly available transcriptome sequencing (RNA-Seq) data. Our findings revealed the underlying RNA virus communities in a wide range of insects, representing a comprehensive study investigating the unexplored RNA viruses in insects, as well as their evolutionary and ecological implications.

RESULTS

Abundant and divergent RNA viruses identified in insects. We performed the virus discovery analyses on a curated collection of RNA-Seq data available publicly in databases. The data set contained 664 species spanning 32 orders of insects (see Data Set S1 in the supplemental material). The assembled contigs were subjected to BLAST

analyses against the reference viral databases. This resulted in ~16,990 virus-associated contigs from 648 insect species, of which 89.35% were associated with two or more virus species (Data Set S1). To elucidate the evolutionary aspect of the newly identified viruses, we extracted the putative viral contigs that had significant matches to the highly conserved viral RNA-dependent RNA polymerase (RdRp) domain and reconstructed the phylogenetic trees together with the recognized taxa of RNA viruses proposed by the International Committee on Taxonomy of Viruses (ICTV [https://talk.ictvonline.org/taxonomy/vmr/]), as well as some unclassified RNA viruses occupying important phylogenetic positions. A total of 1,213 contigs representing complete or partial viral genomes were recovered from the RNA-Seq data for 484 insect species (484/664, 72.89%) (see Fig. S1a in the supplemental material; see also Data Set S1). Phylogenetic analysis demonstrated that these putative RNA viruses fell within or in close proximity to a wide range of recognized viral taxa, including 40 viral families and 2 unclassified genera, as well as newly described RNA virus groups, such as *Qinviridae* and *Chuviridae* (Data Set S1). Nevertheless, BLAST analysis and pairwise comparisons revealed that the sequences of most of the newly identified RNA viruses were highly divergent from the previous reported viral sequences: almost 1,194 (91.01%) shared less than 70% amino acid (aa) identity with the most closely related viruses, while 471 (35.93%) shared less than 40% (Data Set S1; see also Fig. S1b).

The current official taxonomy could not recapitulate the extremely diverse RNA virus populations in insects accurately. Indeed, more than 40% of the novel viruses could not be precisely assigned to the current taxa or exhibited low confidence levels. Instead, we reclassified the RNA viruses into 26 clades as described for previous studies (Fig. 1a) (9, 33). These viral clades were abbreviated based on the representative members of each clade (e.g., "Picorna" is an abbreviation representing *Picornavirales*, *Solinviviridae*, and *Caliciviridae*). Although these viruses were highly divergent, their taxonomic status could be deduced based on the panoramic phylogenies. Newly discovered viruses were assigned into 22 viral clades, of which 13 were relatively ubiquitous as they were observed in more than 10 insect orders (Fig. 1b; see also Data Set S1). The others, including one putative RNA phage within the Levi clade, appeared to be infrequent. A clearly separated group which cannot be confidently assigned to any currently recognized taxon was found on the phylogenies (Fig. 1a). The novel clade comprised 36 viruses that were mostly identified in insect orders *Psocodea* (*n* = 21) and *Hemiptera* (*n* = 9). The closest relative of these viruses was *Agaricus bisporus virus 16* (ABV16), a putative four-segmented virus found in mushroom (25.62% to 44.65% amino acid similarity for RdRp segment, Data Set S1) (34). These viruses might represent a novel viral family or order.

The detailed evolutionary landscape was reevaluated by phylogenetic analyses based on all available RNA viruses by clade. As indicated, viruses identified in insects were distributed throughout different phylogenies (Fig. 2; see also Fig. S2 to S6), implying a broader host range and much-more-diverse RNA virus communities in insects. The observed categories of RNA viruses could be comparable to those of the whole invertebrates, although some taxonomic groups appeared to be more highly represented in insects, including the recognized insect-specific iflaviruses, dicistroviruses, bunyaviruses, rhabdoviruses, flaviviruses, and negeviruses. A growing number of novel RNA viruses formed apparent insect-specific lineages within many other clades, such as Hepe-Beny, Orthomyxo, Tombus, Toti, Tymo, Solemo, Narna, Virga, Nido and Partiti (for definitions of the abbreviations of the virus designation that appear here and throughout, see the Fig. 1 legend). It should be noted that closely related species of plant viruses were frequently discovered in insects. These putative plant viruses might be transmitted by insect vectors and were most commonly found in clade Tymo, as well as some new members of *Secoviridae*, *Virgaviridae*, *Partitiviridae*, *Aspiviridae*, *Botourmiaviridae*, and other plant virus species (Data Set S1).

Novel positive-sense single-stranded RNA (+ssRNA) viruses. The picorna-like viruses were significantly overrepresented in insects (Fig. 1b), accounting for up to

a

b
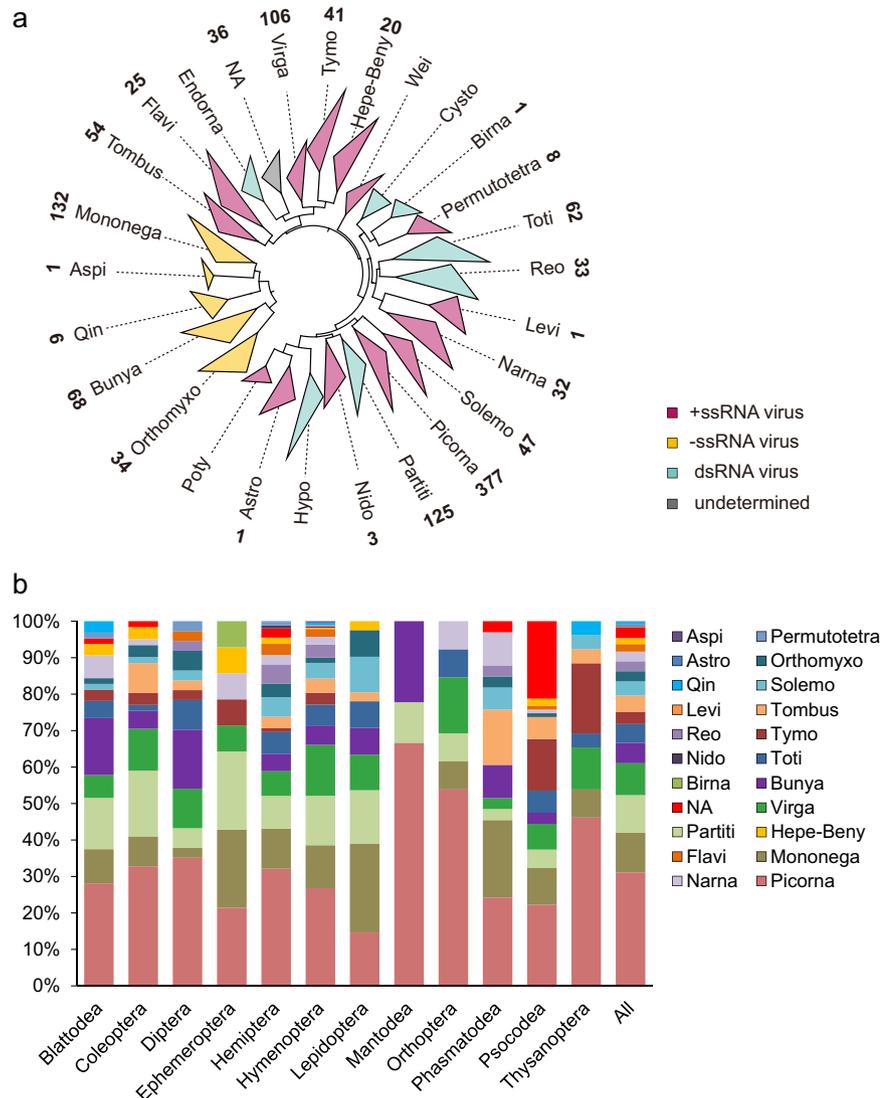
FIG 1 Taxonomic diversity and distribution of RNA viruses in insects. (a) Taxonomic reassessment of the RNA viruses based on RdRp. For each phylogeny, branches are collapsed and are displayed as triangles filled with different colors (see legend). The name of each clade is abbreviated based on the collapsed viral families or orders. The number of viral contigs assigned to each clade is indicated by the Arabic numerals next to the clade name. (b) Percent relative abundance of RNA viruses across each insect order. Those insect orders with too few samples (<10 samples) analyzed in our data set are not listed individually. Astro, *Astroviridae*; Aspi, *Aspiviridae*; Hepe-Beny, *Benyviridae*, *Hepeviridae*, *Alphatetraviridae*, *Togaviridae*, *Bastrovirus*; Birna, *Birnavirdae*; Bunya, *Bunyavirales*, *Arenaviridae*; Cysto, *Cystoviridae*; Endorna, *Endornaviridae*; Flavi, *Flaviviridae*; Hypo, *Hypoviridae*; Levi, *Leviviridae*; Mononega, *Mononegavirales*, *Jingchuvirales*; Narna, *Narnaviridae*, *Botourmiaviridae*; Nido, *Nidovirales*; Partiti, *Partitiviridae*, *Amalgaviridae*, *Picobirnaviridae*; Orthomyxo, *Orthomyxoviridae*; Permutotetra, *Permutotetraviridae*; Poty, *Potyviridae*; Picorna, *Picornavirales*, *Solinviviridaes*, *Caliciviridae*, *Marnaviridae*; Qin, *Qinviridae*; Solemo, *Solemoviridae*, *Luteoviridae*, *Barnaviridae*, *Alvernaviridae*; Reo, *Reoviridae*; Tombus, *Tombusviridae*, *Nodaviridae*, *Sinaivirus*, *Carmotetraviridae*, *Luteoviridae*; Toti, *Totiviridae*, *Chrysoviridae*, *Megabirnaviridae*, *Quadriviridae*, *Botybirnavirus*; Tymo, *Tymovirales*; Virga, *Virgaviridae*, *Togaviridae*, *Bromoviridae*, *Closteroviridae*, *Idaeovirus*; Wei, *Weivirus*; NA, unclassified.

31.08% (377/1,213) of the novel viruses, and were present in 33.73% (224/664) of the insect species. Novel picorna-like viruses were found to diversify into multiple different evolutionary lineages (Fig. S2), in which the insect-associated *Iflaviridae* and *Dicistroviridae* were the most abundant (73.7% of all picorna-like viruses). The family *Secoviridae* was the major group of plant picorna-like viruses (35); which six novel strains were identified with 48.3% to 99.0% aa identity to the most closely related plant virus species. Of those, two close relatives of nepoviruses (93.7% nucleotide [nt] pairwise
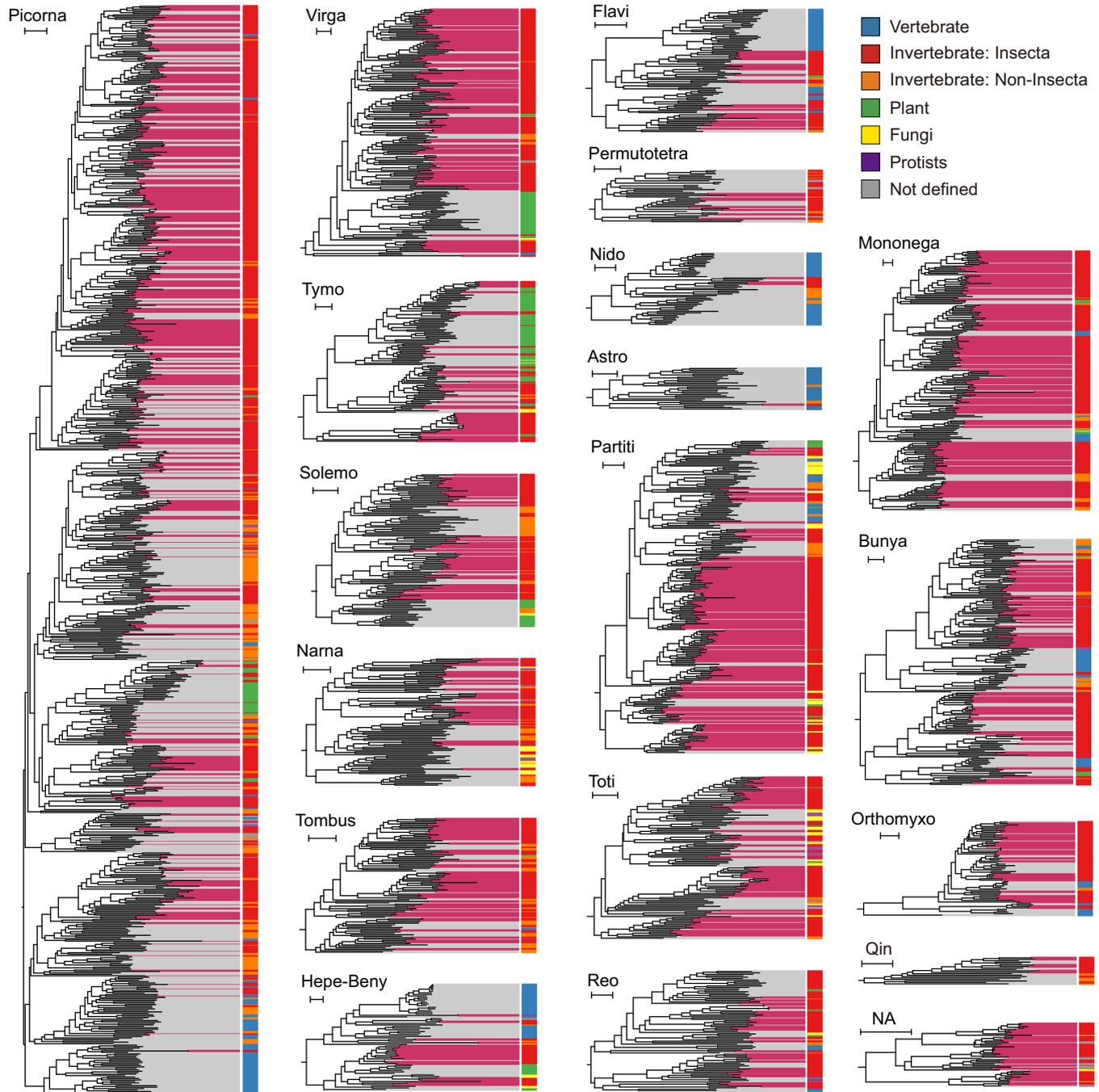
FIG 2 Genetic diversity of RNA viruses in insects. The illustrated phylogenetic trees, which 19 major groups of RNA viruses, were inferred based on RdRp data using the maximum likelihood method and rerooted at midpoint. As indicated, a branch background in red shading represents viruses identified in this study, and a branch background in shading represents those described previously. Host groups are indicated by differently colored bars (see legend). Scale bars represent 0.5 amino acid substitutions per site. Detailed phylogenies are available in Fig. S2 to S6.

identity) were identified in a pollinating bee (*Halictus quadricinctus*) and a predatory wasp (*Crabro peltarius*), respectively, and higher viral relative abundance was identified in *Halictus quadricinctus* (0.17% total reads). It might be expected that *Halictus quadricinctus* acted as a transmission vector for nepoviruses. Other plant virus-like sequences were scattered throughout the trees, such as the strains closely related to the plant-infecting Tomato matilda virus (TmaV) within *Iflaviridae*. In addition, some viruses were found to be evolutionarily related to vertebrate-associated species. For example, we observed a new picornavirus in *Diplonychus rusticus* that best matched Rhimavirus A

(39.5% aa identity), a member of the proposed reptile- and amphibian-specific genus *Rafivirus* within the *Picornaviridae* (36). Another virus that was found in *Hydrochara caraboides* appeared to be closely related to Ampivirus A1 (83.0% aa identity), previously identified in a smooth newt (37). In the unclassified posavirus-like cluster, four novel members were found to be related to the isolates from humans, pigs, rats, and bats, though the identities of the natural hosts have not yet been clarified.

Twenty-five novel flavi-like viruses were identified, including nine in the newly described segmented *Jingmenvirus* (Fig. S3) (38). Novel jingmenviruses clustered together with Wuhan cricket virus (WHCV) and not with the tickborne Alongshan virus (ALSV) and *Toxocara canis* larva agent (TCLA). Notably, a novel jingmenvirus species was present in *Ctenocephalides felis* (cat flea), which might be implicated as a biological vector of Jingmenvirus, as previously described for blood-sucking ticks and mosquitoes (38–40). Another three viruses were clustered with Wenzhou shark flavivirus, representing a disparate evolutionary lineage from *Flavivirus* and *Jingmenvirus*. More viruses fell outside the Flavi-Jingmenvirus, Hepa-Pegivirus, and Pestivirus lineages. These viruses represented a highly divergent subset of unclassified viruses within the Flavi clade.

The newly identified tymovirus-like viruses comprised five major well-supported groups (Fig. S3), among which four groups were affiliated with known plant/fungus virus groups in the order *Tymovirales*. A total of 17 novel strains were recognized as plant viruses. These viruses demonstrated a close relationship with the well-known plant-infecting species, either with a high percentage of identity (70% nt identity with 90% coverage, Data Set S1) or resolved into a single monophyletic clade in the phylogenies. In addition, novel members within the *Maculavirus* were largely hosted by phytophagous or omnivorous insects, representing potential plant viruses (41). Fourteen novel viruses formed a deeply divergent branch distantly related to *Deltaflexiviridae* and were most frequently identified in insect order *Psocodea* (11/14, 78.57%). A BLASTX search revealed that the conserved RdRp domain shared low (26.58% to 34.36%) amino acid similarity with the most closely related viral sequences. The topologies of the phylogenetic trees demonstrated that these viruses belonged to a novel taxonomic group associated with insects, though the host range should be reconfirmed with more data. Within other plant/fungus-associated clades, such as Virga, Solemo, Narna, Tombus, and Hepe-Beny, extremely diverse RNA virus communities were also discovered (Fig. S3). In addition to the putative plant/fungus-infecting strains, most newly identified viruses formed genetically distinct evolutionary lineages, many of which are as yet unclassified. For example, the virgavirus/negevirus-like viruses represented a diverse group of viruses that were associated with insects and other invertebrate animals.

Insect-specific evolutionary lineages were also found in vertebrate-associated virus clades, such as the Nido, Astro, and Hepe clades (Fig. S3). Novel nidoviruses fell into two distinct insect-specific lineages, in which two strains belonged to the previously described *Mesoniviridae*. Another nido-like virus formed a well-supported monophyletic group with Wuhan insect virus 19 and Wuhan nido-like virus 1, representing an unclassified taxonomic group within *Nidovirales*. One distant relative of astrovirus was discovered in a terrestrial insect, *Unaspis euonymi* (*Hemiptera*: *Diaspididae*). The *Unaspis euonymi* astro-like virus clustered with two invertebrate-associated astroviruses (host: *Myriapoda* and *Hirudinea*) and fell in a position basal to the astroviruses of vertebrate on the phylogeny, possibly representing ancient divergence of astroviruses from invertebrates. To our knowledge, the *Unaspis euonymi* astro-like virus was the first astrovirus species identified in insect. One novel bastrovirus strain was also identified in a filter-feeding insect, *Ephemera* sp. *HW-2014*, and shared a moderate level of sequence similarity with two strains isolated from bat and sewage (60.14% and 61.53% aa identity, respectively). Other +ssRNA viruses included nine permutotetraviruses and one levi-like RNA phage (Fig. S3), among which the novel levi-like virus might be associated with the gut microbiota of the host *Lepismachilis y-signata*.

Novel negative-sense single-stranded RNA (−ssRNA) viruses. The mononega-like viruses were the most common −ssRNA viruses identified, presenting in more than 15% insects in this study. Altogether, newly identified mononega-like viruses diversified into nine major evolutionary groups, including six in *Mononegavirales* (Fig. S4). We observed two clearly separate groups within the *Rhabdoviridae*, of which one was related to plant-associated *Nucleorhabdovirus* and the other belonged to vertebrate-associated *Almendravirus*. It may be hypothesized that insect-associated rhabdoviruses arose independently. Other mononegaviruses included members within the *Lispiviridae*, *Xinmoviridae*, and *Artoviridae* lineages and a Borna-like lineage. Thirty-five viruses were assigned to the newly established *Chuviridae* lineage. These chuviruses were divided into three groups with huge differences in their genome architectures, indicating a possible relationship between the genome arrangement and diversity.

Sixty-nine bunya-like viruses were identified, of which most could be classified into three insect-specific groups, namely, *Goukovirus*, *Orthophasmavirus*, and an unclassified sister group to *Phenuiviridae* (Fig. S4). Within the *Feravirus*, two nearly identical strains (>99% nt identity) were identified in *Trialeurodes vaporariorum* and its parasitic counterpart *Encarsia formosa*, respectively, providing evidence of its importance with respect to feeding behavior in virus transmission. A novel tenuivirus strain was identified in *Archaeopsylla erinacei* and was found to be closely related to the Rice stripe tenuivirus (53.8% aa identity). Another strain obtained from a phytophagous insect, *Larinus minutus*, was distantly related to the plant-infecting Citrus concave gum-associated virus. These bunya-like viruses, including the arena-like virus identified in *Heteropsilopus ingenuus*, were assumed to be trisegmented, though the data have shown that some viral genomes lack one or two segments due to inadequate sequencing reads or extremely divergent features.

A remarkable diversity of orthomyxo-like viruses was revealed in insects (Fig. S4). These viruses formed multiple insect-specific lineages in *Quaranjavirus*. In addition, two thogotovirus variants were present in *Plea minutissima* and shared 51.64% amino acid sequence identity. Consistently, these viruses possessed five to six segments, although some genomic segments were missed and some might be redundant. Other −ssRNA viruses included six strains within the newly described bisegmented *Qinvirus* and one within the plant-infecting *Ophiovirus* (Fig. S4). The aspi-like virus was identified in a phytophagous insect, *Panurgus dentipes*, and was assigned to a new species in *Ophiovirus* with an unknown natural plant host.

Novel double-sense double-stranded RNA (dsRNA) viruses. A genetically diverse population was found in clade Partiti with a broad host range, including novel members within *Partitiviridae* and *Amalgaviridae*. More than 125 partitiviruses formed seven major groups, in which 31 novel strains were assigned to the recognized plant/fungus-infecting *Alphapartitivirus* and *Betapartitivirus* (Fig. S5). However, few were identified in genera *Gammapartitivirus* and *Cryspovirus*. Although a novel clade was identified adjacent to the *Deltapartitivirus*, these viruses were distantly related to the phytopathogenic virus species and might represent an invertebrate-specific lineage. Another four groups cannot be classified into any recognized taxa; one of the group was associated with the plant-infecting Maize-associated partiti-like virus. Members of the family *Partitiviridae* were believed to have two genomic segments. However, the capsid-encoded segment was rarely identified within these novel virus groups. It seemed that a much more diverse parititivirus community existed in insects, which might act as major reservoirs and/or transmission vectors for the parititivirus hosted by plants or fungi. In addition, three nonsegmented partiti-like viruses were found in omnivorous insects *Thanasimus formicarius*, *Tetrodontophora bielanensis*, and *Haliplus fluviatilis*. These novel viruses were phylogenetically related to the plant-infecting species within the *Amalgaviridae* and shared similar genome organizations, with two overlapping open reading frames (ORFs).

Three main virus groups within the Toti clade were found in insects (Fig. S5). These viruses showed limited sequence similarity with the recognized totiviruses. For exam-

ple, 20 newly identified toti-like viruses clustered together with Hubei toti-like virus 22 and formed an apparently separate lineage distinct from those of fungus/protozoa giardiaviruses. Another group contained one plant virus strain, Persimmon latent virus, and could be regarded as a sister group to those noninsect viruses. However, novel viruses were occasionally found in the genera *Victorivirus* and *Trichomonasvirus* and one unclassified taxon associated with protozoa. Other toti-like viruses included botybirna- and chryso-like viruses. The botybirnavirus was found in a predatory insect, *Grylloblatta bifratrilecta*, with 30.7% aa identity with the most closely related viruses. The chryso-like viruses were identified in a phytophagous insect, *Psyllopsis fraxinicola*, and a parasitic wasp, *Eurytoma brunniventris*, with a typically four-segmented genome corresponding to those of members of the *Chrysoviridae* family.

Thirty-three reo-like viruses were identified, among which 11 strains were novel members of genus *Orbivirus* and 13 were related to the plant-associated *Fijivirus* (Fig. S5). Other strains were affiliated with different evolutionary lineages of the genera *Coltivirus*, *Phytoreovirus*, *Seadornavirus*, and *Oryzavirus* and four unclassified groups. However, none of the novel reo-like viruses was found in the insect-specific genus *Cypovirus*. Among them, two reo-like viruses were found to be identical to the previously reported strains isolated from *Diaphorina citri* and *Nilaparvata lugens*, respectively (>96% nt identity for all segments), implying wide distribution in the wild. However, most reo-like viruses were highly divergent from all known species; in particular, the multisegmented genome organization made them somewhat more variable.

One birnavirus was identified in a filter-feeding insect, *Isonychia kiangsinensis*. The genome of *Isonychia kiangsinensis* birna-like virus comprised two segments, of which the polymerase-encoded segment best matched the *Drosophila* X virus (42.8% aa identity). The birnavirus was divided into three evolutionary lineages (Fig. S5); one was associated with vertebrate animals, and the other two were associated with invertebrate animals (mostly insects). Hence, a genetically diverse population of birnaviruses might be present in insects and other invertebrate animals.

Great genome flexibility in insect viruses. Though the overall levels of sequence similarity were very low even within each virus clade, some genes were conserved among RNA viruses. The most ubiquitous and conserved motifs were associated with the viral polymerase (Fig. S7a). In general, the key conserved residues arrange themselves in an orderly manner and maintain structural integrity in the RNA polymerase domain, with the exception of few viruses (Fig. S7b). Highly conserved motif C (GDD) appeared to be lost in the polymerase active sites of birnaviruses or to be translocated upstream of motifs A and B in permutotetraviruses and negeviruses (A→B→C to C→A→B). Consensus motifs were also discovered in some essential viral proteins, including the RNA helicase, protease, methyltransferase, and coat proteins (Fig. 3). These viral proteins represent the most highly conserved genome components in the majority of RNA viruses, although some of them are alternatively encoded by different viruses.

Despite the subtle constraints imposed on RNA virus evolution, the insect viruses retained a substantial level of variability in genome size and architectures (Fig. 3; see also Fig. S2 to S6). The narnaviruses had the simplest genome organization of any RNA viruses, yielding to a single open reading frame (ORF) encoding an RNA-dependent RNA polymerase (RdRp). Nevertheless, an additional ORF was found in the complementary strand of the genome for some insect-derived narnaviruses, which encoded a hypothetical protein with unknown function or with no significant homology to known proteins. More differences were identified in virus genomes containing multiple ORFs for nonsegmented viruses or multiple-genome segments for segmented viruses. For example, the virga-like viruses encoded similar panels of nonstructural proteins, while the structural regions were very flexible and characterized by a complex genomic landscape (Fig. 3), which might be associated with a considerable capacity to accommodate different environments and hosts. Within the toti-like viruses, three different genome structures were observed across different evolutionary lineages (Fig. 3), in-
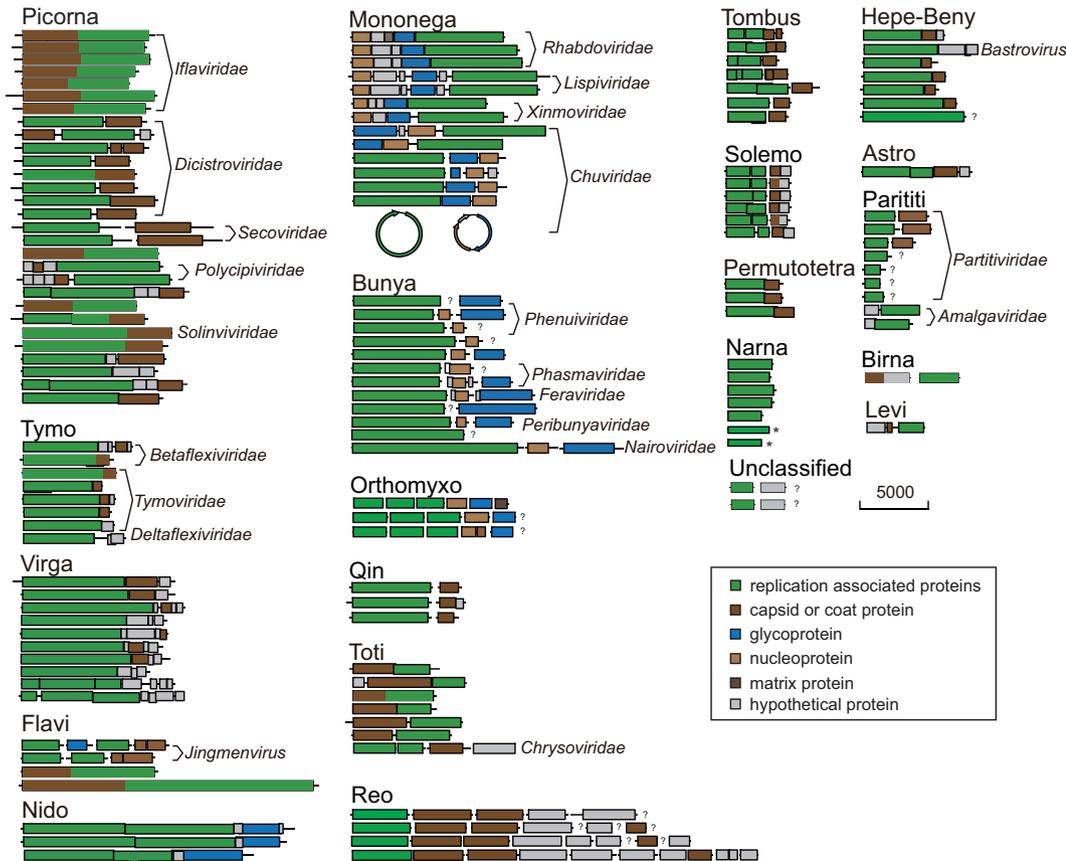
FIG 3 Genome architectures and comparison of representative viruses within major viral clades. The genomes are drawn as boxes and lines approximately to scale, representing open reading frames (ORFs) and noncoding regions, respectively. The predicted homologous genes are shown in colored boxes (see legend). The question marks (?) represent the missing genomic segments, and each asterisk (*) represents a novel hypothetical ORF on the minus strand of the virus genome.

cluding a single large open reading frame (ORF) or discontinuous ORFs characterized by overlapping or being separated. The difference in gene arrangements might reflect an alternative utilization of translation initiation by toti-like and other related viruses, including ribosomal frameshift by those viruses with overlapped ORFs and termination-reinitiation mechanism by those with separated ORFs.

The viral genomic heterogeneity was also enriched through genome rearrangement and segmentation. In picorna-like viruses, one of the most remarkable genetic differences was the gene modules' inversion of the structural and nonstructural protein coding regions among different evolutionary groups (Fig. 3). Within the family *Dicistroviridae*, *Ellipsidion* picorna-like virus 2 and related strains were discovered to undergo an event of genome rearrangement, which likely resulted in the formation of a novel evolutionary lineage descended from *Aparavirus* (Fig. 3). Similarly, the unsegmented mononega-like viruses were characterized by the presence of three hallmark genes encoding the nucleoprotein (N), glycoprotein (G), and large polymerase (L) and were distinguished by a pool of hypothetical proteins with no known orthologues (Fig. 3). Nevertheless, the hallmark genes were not conserved in location in the corresponding genomes, characterized by an N-G-L module for *Mononegavirales* and a G-N-L or L-N-G module for *Jingchuvirales*. Although most mononegaviruses were believed to have a monopartite genome, some chuviruses were characterized by a bi- or trisegmented genome, including a putative circular form (Fig. 3). Genome segmentation was also found in flavi- and solemo-like viruses. In most cases, the segmented viruses were quite distinct from the nonsegmented counterparts. Apart from the conserved NS5 and NS3 segments, the gene repertoires possessed by the segmented jingmenviruses are quite

different from those of the related nonsegmented flavi-like viruses and formed a well-supported monophyletic group. A major phylogenetic discrepancy was also observed between segmented and nonsegmented solemo-like viruses, although the encoded viral proteins were similar. Alternatively, the genomes of insect-associated beny-like viruses were not segmented as seen with the recognized *Benyvirus*. It seems that the segmented viruses might have originated from the primitive nonsegmented ones. The tight evolutionary link between segmented and nonsegmented viruses might represent the potential mechanisms that make possible the genome shuffling in some RNA viruses, such as the mononega- and picorna-like viruses.

DISCUSSION

We conducted a systematic investigation of the RNA viruses in insects, in which abundant and highly diversified RNA virus communities were identified. These RNA viruses were dispersed throughout the phylogenetic trees, and many clades were greatly expanded by the presence of the newly identified viruses (Fig. 2), yielding a considerable fraction within the virus biodiversity. In light of the huge insect populations on the planet, it is conceivable that the level of viral diversity in insects is unprecedented and that a large number of viruses still remain unexplored (15). Based on previous studies and our analysis, most insect-associated viruses are phylogenetically related to other invertebrate viruses, characterized by similar evolutionary tempi and modes (Fig. 2) (42). In general, insect-associated viruses exhibit paraphyletic relationships across the phylogenies of the major virus categories and have diversified into distinct taxonomic groups that comprise more than just several well-studied groups such as *Flaviviridae* (43), *Mononegavirales* (44), and *Bunyavirales* (45). The virus communities possessed by insects appear to be much more genetically diverse than those possessed by other taxonomic kingdoms, including plants and the higher vertebrate animals (9, 10, 46). In the absence of adaptive immunity, insects mainly rely on innate responses for pathogen recognition and clearance (47–49). The lack of adaptive immunity increases the level of tolerance of invasive viruses, ensuring viral survival and accelerated adaptation to the host with little or no fitness cost in insects. Higher viral capacity without an adverse pathological response also results in the insects functioning as important incubators and reservoirs for diverse RNA virus populations, facilitating the emergence of coinfection, recombination, cross-species transmission, and novel viruses.

The remarkable diversity of viral populations might indicate high host specificity of viruses in insects, which would support the idea that viruses are evolutionarily specialized to different ecological niches (11, 13, 50). However, the host barriers are not absolute. A significant risk for viral spillover is associated with trophic interactions, including both antagonism (e.g., herbivory, predation, and parasitism) and mutualism (e.g., pollination), in both directions during contact. The observations in our study demonstrate the possibility of interspecies virus transmission in cases in which predator-prey or host-parasitoid encounters occur, including but not limited to insect-insect interactions. For example, the ectoparasitic mite *Varroa destructor* has facilitated the global spread of deformed wing virus among populations of honeybees, as this method of transmission has allowed circumvention of ecological or evolutionary barriers to infect new hosts (16). Increased direct or indirect contact might confer great opportunities for insect viruses to cross host boundaries. In this respect, the distribution and composition of viruses in insectivorous animals are extensively shaped by insect diet, as indicated by previously reported evidence from bats (13, 51–53). It is commonly thought that hematophagous insects (e.g., mosquitoes, sandflies, and midges) carry a panel of viral pathogens which may be transmitted to humans or animal reservoirs through biting or feeding (54–56). The real risk may be greater than that represented by those examples. Numerous related virus species have been observed in a wide range of insects and can be traced back to the time of host specialization. In most cases, these wild insects do not appear to have direct contact with humans and it remains to be determined whether these viruses possess heterogeneous infectivity and pathogenicity

for humans. Nevertheless, concerns about the potential risk of virus spillover are increasing in accordance with the rapid expansion of human activity and abrupt ecological change (56–58). Unlike animal viruses, most plant viruses are transmitted to their host plants by vectors such as insects, fungi, and nematodes (18, 59). Many insects ingest viruses while feeding on the diseased plants (20, 60), and these viruses have developed the adaptive capacity to manipulate their insect vectors to facilitate their dispersion (18). The indispensable role of intermediate vectors in virus transmission and spread might explain the anomalous number of plant viruses observed in insects.

Analyses of these novel viruses have enabled a comprehensive and integrative reconstruction of the evolutionary history of insect viruses. There appear to be three major scenarios: (i) virus-host codivergence, which is typically characterized by a ladder-like topology in the phylogenies, representing viruses gradually coevolving with their hosts; (ii) virus speciation, in which the viral population undergoes evolutionary branching, resulting in distinct taxonomic groups (it is likely that both internal factors [e.g., genome rearrangement and segmentation] and external factors [e.g., physiological and ecological discrepancies in hosts] play important roles in virus speciation); and (iii) viral spillover, in which the original or ancestral form of the virus can be native or alien to host insects. The former may originate from the pervasive insect-specific viruses (ISV), and virus infection may become established in a new host following their introduction (54). In some circumstances, these viruses achieve productive and persistent infections in the new hosts independently of their original hosts (61, 62). Most of the insect-borne animal viruses, including some vertebrate-specific viruses, would likely fall into this category (43, 63). The prototype of the latter may be the invasive alien viruses introduced by the mechanics of feeding, as often employed in models of predation, herbivory, parasitism, and pollination. This is the choice for the spread of the majority of plant/fungus viruses (60).

In the present study, we have revealed complex viral communities and high levels of multiple infections in insects which may lead to community interactions with important biological consequences. A number of studies have examined the positive/negative effect of insect-specific viruses (ISV) on arbovirus replication in mosquitoes (54). These results highlight the importance of the identification and characterization of the hidden virus communities in trying to understand insect-associated viral diseases. However, great caution must be taken when characterizing a novel virus and its host range. As expected, the endogenous virus elements (EVEs) derived from RNA viruses are usually recognized in the gene pool of an insect genome (64, 65). We could not rule out the presence of EVEs in the present study, especially when the viral genome assembly is incomplete. The host distribution data could also be strongly influenced by unequal sample sizes. In our data set, fewer than 10 examples of RNA-Seq data were available for 20 insect orders, such as the fleas (*Siphonaptera*), representing a long-neglected ectoparasitic insect pest that entails high risk for zoonotic disease (66, 67). Here, extensive studies would provide more-detailed information about the evolution and ecology of virus infections in insects.

MATERIALS AND METHODS

Data set and virus discovery. RNA-Seq data were retrieved from the NCBI Sequence Read Archive (SRA). The entire data set comprised 665 sets of RNA-seq data released by projects "1K Insect Transcriptome Evolution" (1KITE [http://www.1kite.org]) (68), "The evolution of Endopterygota" (69), "Ephemeroptera Phylogenomics" (70), and "Leptree II: non-ditrysia transcriptomes" (71) and other sequencing data for individual species (see Data Set S1 in the supplemental material). Each RNA-Seq data set was assembled *de novo* using the Trinity program with default parameters. The resulting contigs were first subjected to BLASTX searches against a curated database of the publicly available reference viral proteins with an E value cutoff of 1E−5; the database included all recognized RNA virus species approved by the International Committee on Taxonomy of Viruses (ICTV [https://talk.ictvonline.org/taxonomy/vmr/]). The putative viral contigs were then compared to the entire nonredundant protein database (nr) of NCBI using BLASTX to remove false positives, using the same cutoff E value. Viral contigs with unassembled overlaps were merged using the Seqman program implemented in the Lasergene software package (DNASTAR). Among the putative viral contigs, those with predicted amino acid sequence lengths of >100 aa were retained in the following analysis. To facilitate the viral taxonomic classification, the putative viral contigs were also queried against a special database, with a small subset restricted to the conserved viral

RNA-dependent RNA polymerase (RdRp) domain. To remove the false positives conferred by instrument carryover or cross-contamination, the initial reads in the sequencing library were mapped onto the viral contigs using Bowtie2 software with default parameters (72); highly homologous virus sequences (≥95% nt identity) with extremely low (<1%) mapping rates were considered to have been excluded in the subsequent analysis. Putative RNA viruses assigned a designation based on the corresponding hosts from which the virus-like sequences were identified and the virus categories to which they were assigned, followed by a numeral if two or more variants were identified within the same host.

Virus genome annotation. Contigs were annotated by the use of the RAST (Rapid Annotation using Subsystem Technology) server (73) and the online tool ORF Finder (http://www.bioinformatics.org/sms2/ orf_find.html). Conserved domains within the predicted encoded proteins were identified using CD-search against the Conserved Domain Database (CDD) at NCBI. Sequence logos were drawn using the ggseqlogo package in R (74).

Phylogenetic analysis. Global phylogeny analyses were performed along with analyses of the reference viral genomes proposed by the International Committee on Taxonomy of Viruses (ICTV). To fully characterize the individual viral clades, the predicted amino acid sequences were also queried against the nonredundant protein database (nr) at NCBI. Viral protein sequences with significant E values were retrieved and used for phylogenetic reconstruction. In order to reduce computation time, we used the CD-HIT tool (75) to reduce sequence redundancy while maintaining sufficient virus diversity. Sequences were aligned by the use of MAFFT v7.273 and the L-INS-i algorithm (76) and were manually edited in Mega7 (77). Maximum likelihood phylogenetic trees were constructed with the best-fit substitution model implemented in IQ-TREE v1.6.5 (78). Statistical robustness and reliability were assessed using the ultrafast bootstrap (Ufboot) method with 1,000 replicates and the SH-like approximate likelihood ratio test (SH-aLRT) with 1,000 replicates. Finally, trees were visualized using FigTree v1.4.3 (available at http://tree.bio.ed.ac.uk/software/figtree/) and further edited using the online tool iTOL v4 (https://itol.embl.de/). Improvements in the phylogenies were achieved by removing divergent and ambiguous sequences.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, EPS file, 1.6 MB.
FIG S2, EPS file, 1.3 MB.
FIG S3, EPS file, 1.5 MB.
FIG S4, EPS file, 1.4 MB.
FIG S5, EPS file, 0.9 MB.
FIG S6, EPS file, 0.5 MB.
FIG S7, EPS file, 2.8 MB.
DATA SET S1, XLSX file, 2.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Suttle CA. 2007. Marine viruses–major players in the global ecosystem. Nat Rev Microbiol 5:801–812. https://doi.org/10.1038/nrmicro1750.

2. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. Nature 536:425–430. https://doi.org/10.1038/nature19094.

3. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F. 2009. Laboratory procedures to generate viral metagenomes. Nat Protoc 4:470–483. https://doi.org/10.1038/nprot.2009.10.

4. López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A. 2009. High diversity of the viral community from an Antarctic lake. Science 326:858–861. https://doi.org/10.1126/science.1179287.

5. Gilbert KB, Holcomb EE, Allscheid RL, Carrington JC. 2019. Hiding in plain sight: new virus genomes discovered via a systematic analysis of fungal public transcriptomes. PLoS One 14:e0219207. https://doi.org/10.1371/journal.pone.0219207.

6. Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, Grabe M, Hendrix RW, Girones R, Wang D, Pipas JM. 2011. Raw sewage harbors diverse viral populations. mBio 2:e00180-11. https://doi.org/10.1128/mBio.00180-11.

7. Grybchuk D, Akopyants NS, Kostygov AY, Konovalovas A, Lye LF, Dobson DE, Zangger H, Fasel N, Butenko A, Frolov AO, Votypka J, d'Avila-Levy CM, Kulich P, Moravcova J, Plevka P, Rogozin IB, Serva S, Lukes J, Beverley SM, Yurchenko V. 2018. Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite Leishmania. Proc Natl Acad Sci U S A 115:E506–E515. https://doi.org/10.1073/pnas.1717806115.

8. Mushegian A, Shipunov A, Elena SF. 2016. Changes in the composition of the RNA virome mark evolutionary transitions in green plants. BMC Biol 14:68. https://doi.org/10.1186/s12915-016-0288-8.

9. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ. 2016. Redefining the invertebrate RNA virosphere. Nature 540:539–543. https://doi.org/10.1038/nature20167.

10. Shi M, Lin XD, Chen X, Tian JH, Chen LJ, Li K, Wang W, Eden JS, Shen JJ, Liu L, Holmes EC, Zhang YZ. 2018. The evolutionary history of vertebrate

RNA viruses. Nature 556:197–202. https://doi.org/10.1038/s41586-018-0012-7.

11. Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, Biek R, Streicker DG. 2020. Demographic and environmental drivers of metagenomic viral diversity in vampire bats. Mol Ecol 29:26–39. https://doi.org/10.1111/mec.15250.

12. Wille M, Eden J-S, Shi M, Klaassen M, Hurt AC, Holmes EC. 2018. Virus-virus interactions and host ecology are associated with RNA virome structure in wild birds. Mol Ecol 27:5263–5278. https://doi.org/10.1111/mec.14918.

13. Shi C, Beller L, Deboutte W, Yinda KC, Delang L, Vega-Rua A, Failloux A-B, Matthijnssens J. 2019. Stable distinct core eukaryotic viromes in different mosquito species from Guadeloupe, using single mosquito viral metagenomics. Microbiome 7:121. https://doi.org/10.1186/s40168-019-0734-2.

14. Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, Jiang J, Yang J, Dong J, Sun L, Zhu Y, Li Y, Zheng D, Zhang C, Su H, Zheng Y, Zhou H, Zhu G, Li H, Chmura A, Yang F, Daszak P, Wang J, Liu Q, Jin Q. 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. Microbiome 6:178. https://doi.org/10.1186/s40168-018-0554-9.

15. May RM. 1988. How many species are there on Earth? Science 241:1441–1449. https://doi.org/10.1126/science.241.4872.1441.

16. Wilfert L, Long G, Leggett HC, Schmid-Hempel P, Butlin R, Martin SJ, Boots M. 2016. Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. Science 351:594–597. https://doi.org/10.1126/science.aac9976.

17. Manley R, Boots M, Wilfert L. 2015. Emerging viral disease risk to pollinating insects: ecological, evolutionary and anthropogenic factors. J Appl Ecol 52:331–340. https://doi.org/10.1111/1365-2664.12385.

18. Eigenbrode SD, Bosque-Pérez NA, Davis TS. 2018. Insect-borne plant pathogens and their vectors: ecology, evolution, and complex interactions. Annu Rev Entomol 63:169–191. https://doi.org/10.1146/annurev-ento-020117-043119.

19. Liu S, Xie J, Cheng J, Li B, Chen T, Fu Y, Li G, Wang M, Jin H, Wan H, Jiang D. 2016. Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. Proc Natl Acad Sci U S A 113:12803–12808. https://doi.org/10.1073/pnas.1608013113.

20. Blanc S, Gutierrez S. 2015. The specifics of vector transmission of arboviruses of vertebrates and plants. Curr Opin Virol 15:27–33. https://doi.org/10.1016/j.coviro.2015.07.003.

21. Hogenhout SA, Ammar el D, Whitfield AE, Redinbaugh MG. 2008. Insect vector interactions with persistently transmitted viruses. Annu Rev Phytopathol 46:327–359. https://doi.org/10.1146/annurev.phyto.022508.092135.

22. Paixao ES, Teixeira MG, Rodrigues LC. 2018. Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases. BMJ Glob Health 3:e000530. https://doi.org/10.1136/bmjgh-2017-000530.

23. Musso D, Rodriguez-Morales AJ, Levi JE, Cao-Lormeau VM, Gubler DJ. 2018. Unexpected outbreaks of arbovirus infections: lessons learned from the Pacific and tropical America. Lancet Infect Dis 18:e355–e361. https://doi.org/10.1016/S1473-3099(18)30269-X.

24. Mellor PS. 2000. Replication of arboviruses in insect vectors. J Comp Pathol 123:231–247. https://doi.org/10.1053/jcpa.2000.0434.

25. Theze J, Bezier A, Periquet G, Drezen JM, Herniou EA. 2011. Paleozoic origin of insect large dsDNA viruses. Proc Natl Acad Sci U S A 108:15931–15935. https://doi.org/10.1073/pnas.1105580108.

26. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui J-M, Bayne EH, Longdon B, Buck AH, Lazzaro BP, Akorli J, Haddrill PR, Obbard DJ. 2015. The discovery, distribution, and evolution of viruses associated with Drosophila melanogaster. PLoS Biol 13:e1002210. https://doi.org/10.1371/journal.pbio.1002210.

27. Belda E, Nanfack-Minkeu F, Eiglmeier K, Carissimo G, Holm I, Diallo M, Diallo D, Vantaux A, Kim S, Sharakhov IV, Vernick KD. 2019. De novo profiling of RNA viruses in Anopheles malaria vector mosquitoes from forest ecological zones in Senegal and Cambodia. BMC Genomics 20:664. https://doi.org/10.1186/s12864-019-6034-1.

28. Sadeghi M, Altan E, Deng X, Barker CM, Fang Y, Coffey LL, Delwart E. 2018. Virome of >12 thousand Culex mosquitoes from throughout California. Virology 523:74–88. https://doi.org/10.1016/j.virol.2018.07.029.

29. Medd NC, Fellous S, Waldron FM, Xuereb A, Nakai M, Cross JV, Obbard DJ. 2018. The virome of Drosophila suzukii, an invasive pest of soft fruit. Virus Evol 4:vey009. https://doi.org/10.1093/ve/vey009.

30. Remnant EJ, Shi M, Buchmann G, Blacquiere T, Holmes EC, Beekman M, Ashe A. 2017. A diverse range of novel RNA viruses in geographically distinct honey bee populations. J Virol 91:e00158-17. https://doi.org/10.1128/JVI.00158-17.

31. Roberts JMK, Anderson DL, Durr PA. 2018. Metagenomic analysis of Varroa-free Australian honey bees (Apis mellifera) shows a diverse Picornavirales virome. J Gen Virol 99:818–826. https://doi.org/10.1099/jgv.0.001073.

32. Kafer S, Paraskevopoulou S, Zirkel F, Wieseke N, Donath A, Petersen M, Jones TC, Liu S, Zhou X, Middendorf M, Junglen S, Misof B, Drosten C. 2019. Re-assessing the diversity of negative strand RNA viruses in insects. PLoS Pathog 15:e1008224. https://doi.org/10.1371/journal.ppat.1008224.

33. Wolf YI, Kazlauskas D, Iranzo J, Lucia-Sanz A, Kuhn JH, Krupovic M, Dolja VV, Koonin EV. 2018. Origins and evolution of the global RNA virome. mBio 9:e02329-18. https://doi.org/10.1128/mBio.02329-18.

34. Deakin G, Dobbs E, Bennett JM, Jones IM, Grogan HM, Burton KS. 2017. Multiple viral infections in Agaricus bisporus - Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. Sci Rep 7:2469. https://doi.org/10.1038/s41598-017-01592-9.

35. Sanfacon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Wetzel T. 2009. Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. Arch Virol 154:899–907. https://doi.org/10.1007/s00705-009-0367-z.

36. Russo AG, Eden JS, Enosi Tuipulotu D, Shi M, Selechnik D, Shine R, Rollins LA, Holmes EC, White PA. 2018. Viral discovery in the invasive Australian cane toad (Rhinella marina) using metatranscriptomic and genomic approaches. J Virol 92:e00768-18. https://doi.org/10.1128/JVI.00768-18.

37. Reuter G, Boros A, Toth Z, Gia Phan T, Delwart E, Pankovics P. 2015. A highly divergent picornavirus in an amphibian, the smooth newt (Lissotriton vulgaris). J Gen Virol 96:2607–2613. https://doi.org/10.1099/vir.0.000198.

38. Qin XC, Shi M, Tian JH, Lin XD, Gao DY, He JR, Wang JB, Li CX, Kang YJ, Yu B, Zhou DJ, Xu J, Plyusnin A, Holmes EC, Zhang YZ. 2014. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. Proc Natl Acad Sci U S A 111:6744–6749. https://doi.org/10.1073/pnas.1324194111.

39. Ladner JT, Wiley MR, Beitzel B, Auguste AJ, Dupuis AP, II, Lindquist ME, Sibley SD, Kota KP, Fetterer D, Eastwood G, Kimmel D, Prieto K, Guzman H, Aliota MT, Reyes D, Brueggemann EE, St John L, Hyeroba D, Lauck M, Friedrich TC, O'Connor DH, Gestole MC, Cazares LH, Popov VL, Castro-Llanos F, Kochel TJ, Kenny T, White B, Ward MD, Loaiza JR, Goldberg TL, Weaver SC, Kramer LD, Tesh RB, Palacios G. 2016. A multicomponent animal virus isolated from mosquitoes. Cell Host Microbe 20:357–367. https://doi.org/10.1016/j.chom.2016.07.011.

40. Wang ZD, Wang B, Wei F, Han SZ, Zhang L, Yang ZT, Yan Y, Lv XL, Li L, Wang SC, Song MX, Zhang HJ, Huang SJ, Chen J, Huang FQ, Li S, Liu HH, Hong J, Jin YL, Wang W, Zhou JY, Liu Q. 2019. A new segmented virus associated with human febrile illness in China. N Engl J Med 380:2116–2125. https://doi.org/10.1056/NEJMoa1805068.

41. Martelli GP, Sabanadzovic S, Abou Ghanem-Sabanadzovic N, Saldarelli P. 2002. Maculavirus, a new genus of plant viruses. Arch Virol 147:1847–1853. https://doi.org/10.1007/s007050200046.

42. Obbard DJ, Dudas G. 2014. The genetics of host-virus coevolution in invertebrates. Curr Opin Virol 8:73–78. https://doi.org/10.1016/j.coviro.2014.07.002.

43. Shi M, Lin XD, Vasilakis N, Tian JH, Li CX, Chen LJ, Eastwood G, Diao XN, Chen MH, Chen X, Qin XC, Widen SG, Wood TG, Tesh RB, Xu J, Holmes EC, Zhang YZ. 2016. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the Flaviviridae and related viruses. J Virol 90:659–669. https://doi.org/10.1128/JVI.02036-15.

44. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife 4:e05378. https://doi.org/10.7554/eLife.05378.

45. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. 2015. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. Proc Natl Acad Sci U S A 112:7536–7541. https://doi.org/10.1073/pnas.1502036112.

46. Dolja VV, Koonin EV. 2011. Common origins and host-dependent diversity of plant and animal viromes. Curr Opin Virol 1:322–331. https://doi.org/10.1016/j.coviro.2011.09.007.

47. Goic B, Stapleford KA, Frangeul L, Doucet AJ, Gausson V, Blanc H, Schemmel-Jofre N, Cristofari G, Lambrechts L, Vignuzzi M, Saleh MC. 2016. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nat Commun 7:12410. https://doi.org/10.1038/ncomms12410.

48. Poirier EZ, Goic B, Tome-Poderti L, Frangeul L, Boussier J, Gausson V, Blanc H, Vallet T, Loyd H, Levi LI, Lanciano S, Baron C, Merkling SH, Lambrechts L, Mirouze M, Carpenter S, Vignuzzi M, Saleh MC. 2018. Dicer-2-dependent generation of viral DNA from defective genomes of RNA viruses modulates antiviral immunity in insects. Cell Host Microbe 23:353–365.e8. https://doi.org/10.1016/j.chom.2018.02.001.

49. Sanchez-Vargas I, Scott JC, Poole-Smith BK, Franz AW, Barbosa-Solomieu V, Wilusz J, Olson KE, Blair CD. 2009. Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. PLoS Pathog 5:e1000299. https://doi.org/10.1371/journal.ppat.1000299.

50. Virgin HW. 2014. The virome in mammalian physiology and disease. Cell 157:142–150. https://doi.org/10.1016/j.cell.2014.02.032.

51. Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, Shi Z. 2012. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. J Virol 86:4620–4630. https://doi.org/10.1128/JVI.06671-11.

52. Bennett AJ, Bushmaker T, Cameron K, Ondzie A, Niama FR, Parra HJ, Mombouli JV, Olson SH, Munster VJ, Goldberg TL. 2019. Diverse RNA viruses of arthropod origin in the blood of fruit bats suggest a link between bat and arthropod viromes. Virology 528:64–72. https://doi.org/10.1016/j.virol.2018.12.009.

53. Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH, Delwart E. 2010. Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. J Virol 84:6955–6965. https://doi.org/10.1128/JVI.00501-10.

54. Halbach R, Junglen S, van Rij RP. 2017. Mosquito-specific and mosquito-borne viruses: evolution, infection, and host defense. Curr Opin Insect Sci 22:16–27. https://doi.org/10.1016/j.cois.2017.05.004.

55. Ayhan N, Charrel RN. 2017. Of phlebotomines (sandflies) and viruses: a comprehensive perspective on a complex situation. Curr Opin Insect Sci 22:117–124. https://doi.org/10.1016/j.cois.2017.05.019.

56. Mayer SV, Tesh RB, Vasilakis N. 2017. The emergence of arthropod-borne viral diseases: a global prospective on dengue, chikungunya and zika fevers. Acta Trop 166:155–163. https://doi.org/10.1016/j.actatropica.2016.11.020.

57. Daszak P, Cunningham AA, Hyatt AD. 2001. Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta Trop 78:103–116. https://doi.org/10.1016/s0001-706x(00)00179-0.

58. Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, Samuel MD. 2002. Climate warming and disease risks for terrestrial and marine biota. Science 296:2158–2162. https://doi.org/10.1126/science.1063699.

59. Schoelz JE, Stewart LR. 2018. The role of viruses in the Phytobiome. Annu Rev Virol 5:93–111. https://doi.org/10.1146/annurev-virology-092917-043421.

60. Gray SM, Banerjee N. 1999. Mechanisms of arthropod transmission of plant and animal viruses. Microbiol Mol Biol Rev 63:128–148. https://doi.org/10.1128/MMBR.63.1.128-148.1999.

61. Novella IS, Clarke DK, Quer J, Duarte EA, Lee CH, Weaver SC, Elena SF, Moya A, Domingo E, Holland JJ. 1995. Extreme fitness differences in mammalian and insect hosts after continuous replication of vesicular stomatitis virus in sandfly cells. J Virol 69:6805–6809. https://doi.org/10.1128/JVI.69.11.6805-6809.1995.

62. Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. 2014. The evolution and genetics of virus host shifts. PLoS Pathog 10:e1004395. https://doi.org/10.1371/journal.ppat.1004395.

63. Longdon B, Murray GG, Palmer WJ, Day JP, Parker DJ, Welch JJ, Obbard DJ, Jiggins FM. 2015. The evolution, diversity, and host associations of rhabdoviruses. Virus Evol 1:vev014. https://doi.org/10.1093/ve/vev014.

64. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet 6:e1001191. https://doi.org/10.1371/journal.pgen.1001191.

65. Holmes EC. 2011. The evolution of endogenous viral elements. Cell Host Microbe 10:368–377. https://doi.org/10.1016/j.chom.2011.09.002.

66. Harvey E, Rose K, Eden JS, Lawrence A, Doggett SL, Holmes EC. 2019. Identification of diverse arthropod associated viruses in native Australian fleas. Virology 535:189–199. https://doi.org/10.1016/j.virol.2019.07.010.

67. Bitam I, Dittmar K, Parola P, Whiting MF, Raoult D. 2010. Fleas and flea-borne diseases. Int J Infect Dis 14:e667–e676. https://doi.org/10.1016/j.ijid.2009.11.011.

68. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspöck U, Aspöck H, Bartel D, Blanke A, Berger S, Böhm A, Buckley TR, Calcott B, Chen J, Friedrich F, Fukui M, Fujita M, Greve C, Grobe P, Gu S, Huang Y, Jermiin LS, Kawahara AY, Krogmann L, Kubiak M, Lanfear R, Letsch H, Li Y, Li Z, Li J, Lu H, Machida R, Mashimo Y, Kapli P, McKenna DD, Meng G, Nakagaki Y, Navarrete-Heredia JL, Ott M, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767. https://doi.org/10.1126/science.1257570.

69. Peters RS, Meusemann K, Petersen M, Mayer C, Wilbrandt J, Ziesmann T, Donath A, Kjer KM, Aspock U, Aspock H, Aberer A, Stamatakis A, Friedrich F, Hunefeld F, Niehuis O, Beutel RG, Misof B. 2014. The evolutionary history of holometabolous insects inferred from transcriptome-based phylogeny and comprehensive morphological data. BMC Evol Biol 14:52. https://doi.org/10.1186/1471-2148-14-52.

70. Miller DB, Bartlett S, Sartori M, Breinholt JW, Ogden TH. 2018. Anchored phylogenomics of burrowing mayflies (Ephemeroptera) and the evolution of tusks. Syst Entomol 43:692–701. https://doi.org/10.1111/syen.12298.

71. Bazinet AL, Mitter KT, Davis DR, Van Nieukerken EJ, Cummings MP, Mitter C. 2017. Phylotranscriptomics resolves ancient divergences in the Lepidoptera. Syst Entomol 42:305–316. https://doi.org/10.1111/syen.12217.

72. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

73. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

74. Wagih O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 33:3645–3647. https://doi.org/10.1093/bioinformatics/btx469.

75. Li W, Fu L, Niu B, Wu S, Wooley J. 2012. Ultrafast clustering algorithms for metagenomic sequence analysis. Brief Bioinform 13:656–668. https://doi.org/10.1093/bib/bbs035.

76. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.

77. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.

78. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.