1 Sequence diversity and evolution of an iflavivirus family associated with ticks

2 Romain Daveu^{a,b}, Caroline Hervet^a, Louane Sigrist^a, Davide Sassera^b, Aaron Jex^{c,d}, Karine Labadie^e,
3 Jean-Marc Aury^f, Olivier Plantard^a, Claude Rispe^{a,*}

4 ^a INRAE, Oniris, BIOEPAR, Nantes, France; romain.daveu@inrae.fr, caroline.hervet@oniris-nantes.fr,
5 louane.sigrist@oniris-nantes.fr, olivier.plantard@oniris-nantes.fr, clauderispe@inrae.fr

6 ^b Department of Biology and Biotechnology “L. Spallanzani”, University of Pavia, Pavia, Italy;
7 davide.sassera@unipv.it

8 ^c Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Parkville,
9 Victoria, 3052, Australia; jex.a@wehi.edu.au

10 ^d Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria, 3010,
11 Australia; jex.a@wehi.edu.au

12 ^e Genoscope, Institut de biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université
13 Paris-Saclay, Evry, France; Karine.Labadie@genoscope.cns.fr

14 ^f Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-
15 Saclay, 91057 Evry, France; jmAury@genoscope.cns.fr

16 * Correspondence: clauderispe@inrae.fr, Tel.: (33) 2 4068 4000

17 **Abstract:** We studied a family of iflaviruses, a group of RNA viruses frequently found in
18 arthropods, focusing on viruses associated with ticks. Our aim was to bring insight on the
19 evolutionary dynamics of this group of viruses, which may interact with the biology of ticks. We
20 explored systematically *de novo* RNA-Seq assemblies available for species of ticks which allowed
21 to identify nine new genomes of iflaviruses. The phylogeny of virus sequences was not congruent
22 with that of the tick hosts, suggesting recurrent host changes across tick genera along evolution. We
23 identified five different variants with a complete or near-complete genome in *Ixodes ricinus*. These
24 sequences were closely related, which allowed a fine-scale estimation of patterns of substitutions:
25 we detected a strong excess of synonymous mutations suggesting evolution under strong positive
26 selection. ISIV, a sequence found in the ISE6 cell line of *Ixodes scapularis*, was unexpectedly near-
27 identical with *I. ricinus* variants, suggesting a contamination of this cell line by *I. ricinus* material.
28 Overall, our work constitutes a step in the understanding of the interactions between this family of
29 viruses and ticks.

30 **Key-words:** Virus - Arthropods - *Ixodes ricinus* - Transcriptome - RNA-Seq

1. Introduction

Ticks are blood-feeding parasites, which makes of them a "hub" for a number of microorganisms (bacterial pathogens and symbionts, viruses, etc.) potentially interacting with ticks and their vertebrate hosts. We will here focus on one of these interactors, RNA viruses, and especially a family of iflaviruses which has recently been shown to represent an occasional partner of ticks. More broadly, RNA viruses have accompanied the evolution of cellular life from its very beginning or even preceded it [1]. Beyond the restricted number of viruses causing diseases to humans or agronomical species, which have concentrated more studies so far, much remains to discover on the impact of viruses on other organisms and on their evolutionary trajectories. The discovery of new RNA viruses has rapidly increased in the recent years, and they have appeared to be far more prevalent and diverse than previously expected, especially in invertebrates, which exhibit a large and diverse "virosphere" [2,3]. High throughput transcriptome sequencing (or RNA-Seq) has proven to be a tool of choice to characterize the genomes of these viruses (see [4], [3], and even in ticks [5], [6], as this technique does not require to target known sequences, while allowing to retrieve contigs that may contain full viral genomes given a sufficient coverage and sufficient prevalence of the virus in the analyzed tissues/individuals. Iflaviruses represent a particular group of RNA viruses: members of Picornavirales, they are non-envelopped and have a single strand positive sense non segmented genome of ~9-11 kb - see [7] for a general description. This family of viruses has been shown to have a privileged association with insects, being detected in a growing number of species of several orders. The symptoms associated with this infection vary from null to developmental anomalies (deformed wing virus), or mortality [7]. In addition, iflaviruses have also been detected in some Chelicerata, especially in parasitic mites (e.g. deformed wing virus, that affects honeybees but are transmitted by mites [8]). Some iflaviruses have also been identified in ticks, for example in the tick *Hyalomma asiaticum* and in pools of different tick species [3]. Recently, a complete genome of an iflavirus has also been sequenced in an endemic Australian species, *Ixodes holocyclus* [9], in a tick species associated with marine birds, *Ixodes uriae* [10], and in the ISE6 cell line of *Ixodes scapularis* [11]. These findings illustrate two different strategies to isolate microorganisms, which each have their specific potential and interests: collecting and sequencing large pools of wild individuals (allowing to obtain sequences even for viruses that have a potentially low prevalence, as in [5] and [9]), or investigating well-characterized cell lines (as in [11]). Tick cell lines are an important tool for characterizing or experimenting on the interactions between microorganisms and mammalian systems [12] and have been screened for the presence of microorganisms, including viruses [13,14].

In this study, we took advantage of a large number of transcriptome sequencings for different tick species, either produced by our group or available in public repositories, to search for candidate sequences of *Iflaviridae*. We identified nine new genomes of iflaviruses in different tick species (or new variants constituting a quasi-species, for *Ixodes ricinus*), and discuss their evolution and phylogenetic relationships.

2. Materials and Methods

2.1 Search of Iflavirus sequences: We first searched sequences matching iflaviruses in all available public transcriptome assemblies for ticks (available from TSA, Genbank). We searched assemblies (TSA files) rather than raw read (SRA database) because i) an exhaustive search of raw reads would have implied re-assembling de novo all RNA-Seq data sets for all tick species, which represent a considerable volume of data ii) most RNA-Seq projects for ticks have been assembled *de novo* and the assemblies are available in TSA iii) virus sequences are not filtered out from *de novo* assemblies by Genbank, in our experience iv) searches of unfiltered *de novo* assemblies using raw reads, which our group produced recently for 27 different tick species [15] gave the same results as a search to

TSA assemblies, regarding positive or negative detection of iflaviruses. Overall, this suggests an absence of filtering of virus sequences from the public transcriptome assemblies, and that the strategy we used was efficient to detect iflaviruses. For each transcriptome assembly, we searched matches by using the iflavirus described from a cell line of *I. scapularis* (ISIV genome), and doing a tblastn between the poly-protein sequence (complete genome) as a query and RNA-Seq contigs as targets. This search was started with *I. ricinus*, for which a large amount of transcriptomic resources has been generated (with diverse strains and origins of samples), and then extended to other species of ticks. Some of the published assemblies for different species of the genus *Ixodes* had been produced by our research group for a phylogenetic study of hard ticks (Charrier et al., 2019). Our group has additionally recently produced three new transcriptomes for *I. ricinus* (Table 1), with specific objectives (transcriptomic study of the synganglion, transcriptome of whole nymphs, and expression profiling of different organs during feeding). These three transcriptomes have not yet been published as the three studies still need to be finalized, and will be described in upcoming manuscripts, but we used them to complete the search of iflavirus genomes with the same methods.

2.2 Phylogeny methods: The phylogenetic analysis included all of the iflavirus genome sequences associated with ticks, as identified above. Based on the recent phylogenetic study of iflaviruses [16], we identified two relevant outgroups, which were also included. The first was an iflavirus associated with *Tetranychus truncatus* (representing the closest outgroup to tick iflaviruses) while the second was associated with *Apis mellifera* (Deformed wing virus, or Dfw) - this was a more distant outgroup, but still pertaining to the same cluster (Cluster 1) in the phylogeny of Kobayashi et al. [16]. The alignment of all protein sequences was done with MUSCLE in MegaX [17]. This alignment was filtered with Gblocks [18], which excluded poorly aligned regions and gaps. A phylogenetic ML tree was obtained with IQ-TREE [19]. The best model of substitution was determined with Model Finder [20], then we assessed branch support with 1000 ultrafast bootstrap replication [21]. A graphical edition of the consensus ML tree was performed with ITOL [22].

2.3 Test of congruence between host and virus phylogenies

To test the congruence between virus and host phylogenies, we used the cophylogeny testing tool Jane 4 [23]. As three of the iflavirus genomes were obtained from pools of tick hosts, from different genera, we could not include these three genomes and made this analysis with a subset of the tree, containing all other sequences. We also collapsed virus sequences that were nearly identical in the virus phylogeny, since they can be considered to be variants and belong to the same species. This concerned two iflaviruses associated with *I. holocyclus*, but also a group of six sequences, five being found in *I. ricinus* and the latter in a cell line of *I. scapularis* (see Results): the name given to this group was IricIV-ISIV.

2.4 Estimation of genetic distances and evolutionary rates: We estimated genetic distances and evolutionary rates within two ensembles of sequences (respectively ISIV + *I. ricinus* variants, and two variants associated to *I. holocyclus*) that grouped closely in the phylogenetic analyses. We first estimated genetic distances (number of substitutions per site) both at the nucleotide level and at the amino acid level, using the Maximum Composite Likelihood model [24] and the Poisson correction model [25] respectively, with the complete deletion option (all gaps excluded). This analysis was performed in MEGA X [17]. Then, we estimated ratios of non-synonymous and synonymous substitutions (dN/dS) with Codeml [26]: for this we used the one-ratio model on a sub-tree comprising the ISIV sequence plus five variants found in *I. ricinus* and a pairwise estimate of dN/dS for the two variants found in *I. holocyclus*.

121 3. Results

122 **3.1 Iflavirus sequences identified.** Five variants of an iflavirus with a complete or near-complete
 123 genome could be found in different transcriptomes of *Ixodes ricinus* - see Table 1 for details. One
 124 variant was found in an assembled transcriptome of synganglions from unfed wild adult ticks (Rispe
 125 et al. in prep.), the corresponding sequence being named IricIV-1. While there was read support for
 126 the corresponding contig in eight of nine libraries, an overwhelming majority (99%) of these reads
 127 came from just one library. A second complete sequence (IricIV-2) could be retrieved from a newly
 128 obtained transcriptome of different tissues (Daveu et al. in prep.), from *I. ricinus* individuals from
 129 the lab strain from the University of Neuchâtel, Switzerland. We also explored published assemblies
 130 of *I. ricinus* transcriptomes and detected a third complete iflavirus genome (IricIV-3) in the recently
 131 published GIDG data set (TSA accession from NCBI). Last, we identified two iflavirus genomes in
 132 transcriptomes of pooled whole bodies of nymphs (Rispe et al., in prep.) collected on the
 133 vegetation, respectively in Switzerland (IricIV-4) and north-east France (IricIV-5). All viral
 134 genomes obtained in *I. ricinus* were highly similar to the ISIV sequence obtained from a cell line of
 135 *I. scapularis* (~98%, amino acid identity based on tblastn matches, while higher variation was
 136 detected at the nucleotide level, as shown below). All other data sets of reads from *I. ricinus* either
 137 showed absence of matches, or in one case, small fragmented contigs (GCJO data set) which we
 138 will not further consider here. Although the ISIV sequence was found in a cell line transcriptome of
 139 *I. scapularis*, there was no match with this sequence in *de novo* assemblies obtained for three other
 140 independently sequenced transcriptomes of *I. scapularis* (two of them at least obtained from large
 141 collections of wild ticks, including a total of ~200 males and females from three locations for the
 142 GGIX01 accession number). Then, we also screened published assemblies for other tick species, as
 143 available in Genbank-TSA: this search was positive for *I. frontalis* (IfronIV) and *I. vespertilionis*
 144 (IvespIV) but in both cases the retrieved sequences were incomplete (5' partial), and the obtained
 145 ORF comprised respectively 2437 and 1945 amino acids (representing ~81% and ~65% of the
 146 complete genome respectively). Additionally, a 5' partial sequence was also found in a
 147 transcriptome assembly of *I. holocyclus* (2039 amino acids, IhIV-2). This sequence (and the
 148 corresponding samples and sequencing) was different from the first iflavirus genome (IhIV)
 149 identified for that tick species. We also found a matching contig sequence in an assembled
 150 transcriptome (GFGI01 TSA assembly) of the tick *Hyalomma dromaderii*. The match was in two
 151 frames suggesting a likely frameshift, probably due to sequence errors in the reads, or in the
 152 assembly. We tentatively corrected the frameshift manually (with one base deletion) and obtained a
 153 complete iflavirus genome sequence for this species. All other assemblies, including several genera
 154 of hard and soft ticks proved negative for this search. Finally, other sequences were found after a
 155 blast (blastp) to the Genbank nucleotide database (nr): this allowed to include sequences of
 156 iflaviruses associated respectively with *Amblyomma americanum*, *Haemaphysalis flava*,
 157 *Hamephysalis asiaticum*, *Ixodes holocyclus*, *Ixodes uriae*, or with pools of tick species (Table 1).

158 **3.2 Phylogeny.** To obtain a comprehensive phylogeny of iflavirus genome sequences associated
 159 with ticks, we integrated all the available genome sequences (as detailed in Table 1), comprising
 160 sequences already described [3,9,11,16] and the nine sequences newly identified in the present
 161 study (Fig. 1). After filtering of regions that aligned poorly and of gaps, the alignment comprised
 162 1149 amino acid sites. The best-fit model (BIC criteria) was LG+G4. The ML tree (consensus from
 163 1000 bootstrap trees) showed that the five variants found in *I. ricinus* and the ISIV sequence closely
 164 grouped together - see below for details on genetic distances. This was also the case of the two
 165 variants of *I. holocyclus*. Thus two groups, respectively ISIV+*I. ricinus* variants and *I. holocyclus*
 166 variants, form a 'quasi-species'. Another iflavirus sequence, found in *I. vespertilionis*, formed a
 167 sister group with the two groups above. Two other iflavirus genomes, respectively found in *I.*
 168 *frontalis* and *H. dromaderii*, grouped with Ht-V1, a sequence obtained in the study of Shi et al. [3].

169 The latter sequence was obtained from a pool of tick species which contained several species of the
170 Metastriata (non-*Ixodes* hard ticks) or soft ticks species (Argasidae), but no species of the genus
171 *Ixodes*. Of note, a sequence found in *A. americanum* in the study of Tokarz et al. [5] corresponds to
172 an iflavirus although it was named "dicistrovirus" in this manuscript.

173 3.3 Test of cophylogeny

174 The test of cophylogeny with Jane 4 allowed us to retrieve several possible coevolutionary
175 scenarios, which all had the same scoring and differed very little in structure (we represent one of
176 them in Fig. 2). For all scenarios, there was an initial cophylogeny, due to the fact that all iflavirus
177 genomes associated with ticks form a monophyletic clade. But after this ancestral node (following
178 the association of iflaviruses to ticks), all scenarios found several events of incongruence, with four
179 instances of host switch. An additional anomaly was a "failure to diverge" between viruses
180 associated with respectively *I. ricinus* and *I. scapularis* (see Discussion).

181 3.4 Distances among iflavirus genome sequences and rates. For the group comprising the variants
182 found in *Ixodes ricinus* plus the ISIV sequence, the estimated pairwise distances ranged between
183 0.010 and 0.078 at the nucleotide level, suggesting that these sequences are closely related but not
184 identical, whereas amino acid distances were lower, ranging between 0.004 and 0.024 (Table 2).
185 Consistently, the estimated non-synonymous to synonymous ratio was low and well below one
186 (with the one ratio model, dN/dS=0.024). The sequences of the two variants found in *Ixodes*
187 *holocyclus* were extremely close but not identical (105 differences over 6117 nucleotide positions,
188 but only 5 differences at the amino acid level). For this pair of variants, the estimated pairwise
189 distances were respectively 0.018 (nucleotide level) and 0.002 (amino acid level) whereas the
190 estimated ratio of non-synonymous to synonymous ratio was also very low (pairwise ratio,
191 dN/dS=0.014).

192 4. Discussion

193 Communities of microorganisms associated with ticks are attracting a special interest in the context
194 of the growing concern for tick-transmitted diseases [27]. In recent years, there has been a particular
195 interest for viruses associated with ticks, and in particular for the iflavirus family since several tick
196 species appeared to harbour iflavirus-like viral genomes [3,5,9,11,16]. Our study, based on the
197 exploration of assembled transcriptomes, allowed to discover nine new iflavirus genome sequences
198 in ticks, and to perform a phylogenetic study that includes both these new genome sequences and
199 previously published sequences. Our phylogenetic analysis showed that the iflavirus genome
200 sequences associated to ticks closely grouped together and formed a monophyletic clade. This
201 suggests an ancient association between this virus sub-group and ticks, which could mean relatively
202 infrequent host-switches of the virus between major groups of arthropods. Of note however, two
203 iflavirus-like genome sequences have been recently described for the *Antricola* genus [28], a genus
204 of neotropical soft ticks which have a peculiar biology, being associated with hot bat caves and
205 partially feeding on bat guano instead of an exclusive vertebrate blood diet. These two sequences
206 (not included in our phylogeny) grouped closely with insect iflaviruses and not with sequences
207 associated with ticks, showing that there has been more than one infection of ticks by this family of
208 viruses. Within ticks, the virus phylogeny and the tick phylogeny were not congruent, i.e. there was
209 no strict co-cladogenesis (Fig. 1 and Fig. 2). On the one hand, a monophyletic group comprised
210 sequences found in *I. ricinus* (or *I. scapularis* - see below), *I. holocyclus* and *I. vespertilionis*,
211 suggesting that they form an "*Ixodes*" subclade within iflavirus associated to ticks. However, the
212 virus sequence found in *I. frontalis* did not group with this ensemble, and was closer to iflavirus
213 genomes found in ticks from other genera. Also, the closer grouping of *I. holocyclus* and *I. ricinus*

was inconsistent with the phylogeny of this genus [15,29], since *I. holocyclus*, a species belonging to the Australasian subgroup of *Ixodes* is phylogenetically distant from *I. ricinus* and *I. vespertilionis*. Although the data for iflaviruses in genera other than *Ixodes* is still incomplete or imprecise (e.g. several sequences came from assembled transcriptomes based on pooled species of ticks), we finally note that the sequences associated with two species of the genus *Hyalomma* did not group together. In addition the formal test of cophylogeny (Fig. 2) on a subset of virus sequences (some viruses could not be included in the test, since they could not be assigned to a precise host) suggests several events of switching of the virus between ticks species or genera over evolutionary time. A more precise picture of this dynamics will require a denser data set including more iflavivirus genomes, and a larger sample of tick species.

RNA viruses have relatively fast evolving sequences, which was also observed in our data set, as we compared the virus genomes associated with different tick species. There was however a striking exception, concerning the sequence found in a cell line of *I. scapularis* (ISIV), which formed a tight cluster in the ML phylogenetic tree with five different variants identified in *Ixodes ricinus*. In fact these sequences were nearly identical at the amino acid level (>98% pairwise identity of protein sequences). Therefore, ISIV and the *I. ricinus* iflavivirus sequences form together a 'quasi-species', a puzzling result given that the tick host species *I. scapularis* and *I. ricinus* have evolved on two different continents (respectively north-east America and Europe) and are separated by several million years of divergence. The geographic distance and the scarcity of animal (birds would be the only shared potential tick hosts) movements between Western Europe and North America leaves little chance of a recent natural inter-species contamination through shared tick hosts. We thus consider a different scenario to explain this anomaly: ISIV was found in the transcriptome of a cell line of *I. scapularis* (ISE6), but we failed to detect it in any other assembled transcriptome of this species, including a data set comprising as many as 200 adult ticks from several locations [30]. On top of that, another study assessed the presence of viral sequences in *I. scapularis* after surveying several pools of wild samples (a total of >1100 individuals, combining nymphs and adults) from different locations - New York and Connecticut, USA - but did not report any iflavivirus-like sequence [5]. Of note, with the same methods, this study did identify an iflavivirus genome in *Amblyomma americanum* (although it was named "dicistrovirus", this sequence groups with iflavirus). By contrast, the five *I. ricinus* variants found in the present study were obtained for several independent samples of wild ticks (or from one lab strain initially derived from wild ticks), which strongly supports the fact that the *I. ricinus* iflavivirus variants newly discovered in our study correspond to strains of iflavivirus that naturally infect this species. Overall, this apparent absence of the virus (or at least of an ISIV-like sequence) in wild individuals of *I. scapularis* combined with near-identity between ISIV and the sequences associated with *I. ricinus* could be better explained by a contamination of the *I. scapularis* ISE6 cell line with a virus associated in the field with *I. ricinus*. Such contamination could have been mediated by the establishment or maintenance of cell lines of the two species in a same lab, a possibility previously discussed for other viruses and cell lines of *I. scapularis* in the study of Alberdi et al. [13].

Much remains to know on the evolutionary dynamics of the association between iflaviruses and tick species. This first concerns the modes of transmission of the virus, and the possibility of transmission among ticks of different species (or genera) that could occur through co-feeding on the same host. An important missing piece of information in the understanding of this system is also the biological effect(s) of iflaviruses on their tick hosts, as there has not been report, to our knowledge, of symptoms in ticks infected by the virus. An element to evaluate the short-term aspects of the evolution of the virus is the level of intra-specific variability (variability between variants of the viral genome found in one tick species). Our study allows, for the first time, to evaluate this variability, since we could obtain and analyse a total of six variants for the group ISIV+ *I. ricinus*,

and also two variants for *I. holocyclus*. The observed variation among strains within each of these groups was mostly constituted of synonymous mutations, and we found evidence of strong purifying selection acting between these variants. This means that at least for short time paces of evolution (within a host tick species), the rate of evolution of the protein sequence of this virus is low.

The extant data is still too scarce to determine the prevalence of the virus among tick species: negative results in several RNA-Seq data for *I. ricinus*, whereas other transcriptomes of this species allowed to detect five different iflavirus genomes, which suggests that the virus is not consistently present in tick populations (and perhaps even that its prevalence is low). The negative results obtained in most species (i.e. all transcriptomes of twelve species of the Metastriata and three sp. of *Ornithodoros*) should therefore not be taken as evidence of absence of this family of virus from these species. Our study, which surveyed different RNA-Seq data sets obtained with very heterogeneous sources of materials (cell lines, lab strains, wild ticks) allowed however to significantly enrich the range of tick species with a complete iflavirus genome. This comforts the scenario of an ancient association and relatively widespread presence of iflavirus genomes in tick species. We suggest that a more systematic use of RNA-Seq, based on large pools of wild individuals, would maximize chances of detecting that type of sequences even in case of low prevalence - see for example [31]. More generally, this would enhance progress in the description of the virome associated with ticks, comprising viruses pathogenic for their vertebrate hosts, including humans [32].

Supplemental material : the file suppl_rispe.txt contains the Genbank format sequences of the newly identified iflavirus genomes.

Author Contributions: Conceptualization and Methodology: C.R. and O.P.; Ressources and Funding acquisition: C.H., D.S., A.J., J.-M.A., K.L., O.P.; Formal analysis and Data curation: C.R., R.D., L.S.; Writing: C.R., O.P. and R.D. All authors have read and agreed to this version of the manuscript.

Acknowledgements: we thank O. Rais, M. Voordouw and Franck Boué who provided part of the material used for transcriptome sequencing of *I. ricinus*. We acknowledge the support of the Genotoul bioinformatics platform Toulouse Midi-Pyrénées (Bioinfo Genotoul) for bioinformatic analyses. This work was supported by the Genoscope, the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) and France Génomique (ANR-10-INBS-09-08), by the Italian Ministry of Education, University and Research (MIUR): Dipartimenti di Eccellenza Program (2018–2022) - Dept. of Biology and Biotechnology "L. Spallanzani", University of Pavia (to D.S.), by a grant from the Institut Carnot France Futur Elevage (XENOBOTICK project). Thanks to Sara Moutailler (ANSES) for helpful comments on the manuscript.

Conflict of interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Table 1. List of iflavirus sequences found in tick transcriptomes (with tblastn or blastp, using the ISIV genome, associated with the ISE6 cell line of *Ixodes scapularis* as a query). Columns: Species, Tissues (HEM: haematocytes, MG: midgut, MT: Malpighian tubules, OV: ovaries, SG: salivary glands, SYN: synganglion, WB: whole bodies) followed by details on the stages or conditions between parentheses, when available, Location of the sampling (or source of the strains), Accessions: Genbank accession and if available, related TSA or BioProject accession between parentheses, Publication (or authors of the sequences), Percent identity with ISIV -% id at the amino acid level of the first hsp (tblastn)- and range of the match, which indicates genome completion. Lines in bold correspond to the nine iflavirus genome sequences newly discovered in the present study.

Species	Tissues and conditions	Location	Accessions	Publication	Identity	Match range(s)	Virus name
<i>Amblyomma americanum</i>	WB (wild questing ticks)	NY and Connecticut, USA	ASU47553.1	Tokarz et al. 2018	65.8%	1142-2838	Lone star tick dicistrovirus
<i>Haemaphysalis flava</i>	WB	Japan	BBK20270.1	Kobayashi et al. 2020	45.6%	63-2937	HtIFV
<i>Hyalomma asiaticum</i>	WB	China	APG77501.1	Shi et al. 2016	48.4%	379-2990	Bole Hyalomma asiaticum
<i>Hyalomma dromaderii</i>	SG, wild ticks	Tunisia	BK012003 (GFGI01)	Bensaoud et al. 2018	40.0%	1-1880 and 2000-2989	HdromIV
<i>Ixodes frontalis</i>	WB	Carquefou, France	MT008333 (PRJNA528282)	Charrier et al. 2019	41.1%	560-2989	IfronV
<i>Ixodes holocyclus</i>	SG, WB, MG	QLD and NSW, Australia	KY020412.1 (GIBQ01)	O'Brien et al. 2018	62.2%	1-2989	IhIV
<i>Ixodes holocyclus</i>	WB	NSW, Australia	MT008332 (PRJNA528282)	Charrier et al. 2019	65.4%	913-2989	IhIV-2
<i>Ixodes ricinus</i>	SYN (wild unfed ticks)	Chizé, France	MT008330	Rispe et al (unpub.)	97.7%	1-2991	IricIV-1
<i>Ixodes ricinus</i>	SG, OV, MT (feeding adult females)	Neuchâtel (Switzerland), lab strain	MT008331	Daveu et al (unpub.)	98.4%	1-2991	IricIV-2
<i>Ixodes ricinus</i>	WB (various stages)	Czech Republic	BK012002 (GIDG01012278)	Vechtova et al. (unpub.)	98.1%	1-2991	IricIV-3
<i>Ixodes ricinus</i>	WB (wild nymphs)	Switzerland	MT050463	Rispe et al (unpub.)	98.3%	6-2991	IricIV-4
<i>Ixodes ricinus</i>	WB (wild nymphs)	Nancy, France	MT050464	Rispe et al (unpub.)	98.2%	1-2991	IricIV-5
<i>Ixodes scapularis</i>	cell line ISE6	-	BBD75427.1	Nakao et al. 2017	100.0%	1-2991	ISIV
<i>Ixodes uriae</i>	WB	Antartic peninsula	QIS88066	Wille et al. 2020	42.0%	383-2991	Gerbovich virus
<i>Ixodes versperilionis</i>	WB	Maine-et-Loire, France	MT008334 (PRJNA528282)	Charrier et al. 2019	68.2%	1016-2989	IvespIV
pool of tick species	WB	China	YP_009336552.1	Shi et al. 2016	38.6%	381-2991	Ht-V1
pool of tick species	WB	China	YP_009336542.1	Shi et al. 2016	44.3%	391-2991	Ht-V2
pool of tick species	WB	China	YP_009336533.1	Shi et al. 2016	46.1%	63-2989	Ht-V3

310 **Table 2.** Pairwise distances among complete genomes of iflavirus associated with *Ixodes ricinus*
 311 (five variants, IricIV-1 to 5, first identified in the present study) or the *I. scapularis* cell line ISE6
 312 (ISIV). Number of base substitutions per site (Maximum composite likelihood model, 8976
 313 positions) followed by number of amino acid substitutions per site (Poisson correction model, 2991
 314 positions). Analyses were conducted with MegaX.

	IricIV-1	IricIV-2	IricIV-3	IricIV-4	IricIV-5	ISIV
IricIV-1						
IricIV-2	0.059 / 0.015					
IricIV-3	0.078 / 0.020	0.064 / 0.015				
IricIV-4	0.060 / 0.016	0.010 / 0.004	0.066 / 0.016			
IricIV-5	0.076 / 0.019	0.065 / 0.015	0.028 / 0.006	0.066 / 0.017		
ISIV	0.077 / 0.024	0.049 / 0.016	0.065 / 0.019	0.050 / 0.017	0.062 / 0.019	

Figure legends:

Figure 1: **A**, Maximum Likelihood phylogenetic tree of tick-associated iflaviruses. The tree was based on the amino acid sequence of iflavirus sequences found in the transcriptomes of different hard tick species (Acari;Parasitiform;Ixodida;Ixodidae), and rooted with two outgroups, a honeybee iflavirus (Deformed wing virus, or Dfw) and an iflavirus associated with *Tetranychus truncatus* (Acari; Acariform; Prostigmata). Host taxon is indicated between brackets. Details related to each taxon (tick host species, source of data, accession, genome completeness) are given in Table 1. Taxon names in bold correspond to the nine iflavirus sequences newly discovered in the present study. Bootstrap support is indicated at the nodes. **B**, Expected topology of the phylogenetic tree of arthropod hosts of iflaviruses included in this study, based on [14] and on the delimitation of the two groups recognized within Ixodidae (Prostriata and Metastrata).

Figure 2: Test of the congruence between the phylogenies of tick-associated iflaviruses and of tick hosts, with Jane 4. This tree contains only part of the sequences analyzed in Fig.1 (three virus genomes found in pools of tick species could not be included). Iric-ISIV represents iflavirus sequences found either in *I. ricinus* or in a cell line of *I. scapularis* (these sequences being nearly identical). Open circles at the nodes indicate co-phylogeny, HS indicates a host-shift, FD a failure to diverge.

Figure 1

335 References

1. Koonin, E.V. On the Origin of Cells and Viruses: Primordial Virus World Scenario. *Ann. N. Y. Acad. Sci.* **2009**, *1178*, 47–64, doi:10.1111/j.1749-6632.2009.04992.x.
2. Li, C.-X.; Shi, M.; Tian, J.-H.; Lin, X.-D.; Kang, Y.-J.; Chen, L.-J.; Qin, X.-C.; Xu, J.; Holmes, E.C.; Zhang, Y.-Z. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* **2015**, *4*, e05378, doi:10.7554/eLife.05378.
3. Shi, M.; Lin, X.-D.; Tian, J.-H.; Chen, L.-J.; Chen, X.; Li, C.-X.; Qin, X.-C.; Li, J.; Cao, J.-P.; Eden, J.-S.; et al. Redefining the invertebrate RNA virosphere. *Nature* **2016**, *540*, 539–543, doi:10.1038/nature20167.
4. Liu, S.; Chen, Y.; Bonning, B.C. RNA virus discovery in insects. *Curr. Opin. Insect Sci.* **2015**, *8*, 54–61, doi:10.1016/j.cois.2014.12.005.
5. Tokarz, R.; Sameroff, S.; Tagliafierro, T.; Jain, K.; Williams, S.H.; Cucura, D.M.; Rochlin, I.; Monzon, J.; Carpi, G.; Tufts, D.; et al. Identification of Novel Viruses in *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* Ticks. *mSphere* **2018**, *3*, doi:10.1128/mSphere.00614-17.
6. Harvey, E.; Rose, K.; Eden, J.-S.; Lo, N.; Abeyasuriya, T.; Shi, M.; Doggett, S.L.; Holmes, E.C. Extensive Diversity of RNA Viruses in Australian Ticks. *J. Virol.* **2019**, *93*, doi:10.1128/JVI.01358-18.
7. Valles, S.M.; Chen, Y.; Firth, A.E.; Guérin, D.M.A.; Hashimoto, Y.; Herrero, S.; de Miranda, J.R.; Ryabov, E.; ICTV Report Consortium ICTV Virus Taxonomy Profile: Iflaviridae. *J. Gen. Virol.* **2017**, *98*, 527–528, doi:10.1099/jgv.0.000757.
8. de Miranda, J.R.; Genersch, E. Deformed wing virus. *J. Invertebr. Pathol.* **2010**, *103*, S48–S61, doi:10.1016/j.jip.2009.06.012.
9. O'Brien, C.A.; Hall-Mendelin, S.; Hobson-Peters, J.; Deliyannis, G.; Allen, A.; Lew-Tabor, A.; Rodriguez-Valle, M.; Barker, D.; Barker, S.C.; Hall, R.A. Discovery of a novel iflavirus sequence in the eastern paralysis tick *Ixodes holocyclus*. *Arch. Virol.* **2018**, *163*, 2451–2457, doi:10.1007/s00705-018-3868-9.
10. Wille, M.; Harvey, E.; Shi, M.; Gonzalez-Acuña, D.; Holmes, E.C.; Hurt, A.C. Sustained RNA virome diversity in Antarctic penguins and their ticks. *ISME J.* **2020**, 1–15, doi:10.1038/s41396-020-0643-1.
11. Nakao, R.; Matsuno, K.; Qiu, Y.; Maruyama, J.; Eguchi, N.; Nao, N.; Kajihara, M.; Yoshii, K.; Sawa, H.; Takada, A.; et al. Putative RNA viral sequences detected in an *Ixodes scapularis*-derived cell line. *Ticks Tick-Borne Dis.* **2017**, *8*, 103–111, doi:10.1016/j.ttbdis.2016.10.005.
12. Munderloh, U.G.; Lynch, M.J.; Herron, M.J.; Palmer, A.T.; Kurtti, T.J.; Nelson, R.D.; Goodman, J.L. Infection of endothelial cells with *Anaplasma marginale* and *A. phagocytophilum*. *Vet. Microbiol.* **2004**, *101*, 53–64, doi:10.1016/j.vetmic.2004.02.011.
13. Alberdi, M.P.; Dalby, M.J.; Rodriguez-Andres, J.; Fazakerley, J.K.; Kohl, A.; Bell-Sakyi, L. Detection and identification of putative bacterial endosymbionts and endogenous viruses in tick cell lines. *Ticks Tick-Borne Dis.* **2012**, *3*, 137–146, doi:10.1016/j.ttbdis.2012.05.002.
14. Bell-Sakyi, L.; Zwegarth, E.; Blouin, E.F.; Gould, E.A.; Jongejans, F. Tick cell lines: tools for tick and tick-borne disease research. *Trends Parasitol.* **2007**, *23*, 450–457, doi:10.1016/j.pt.2007.07.009.
15. Charrier, N.P.; Hermouet, A.; Hervet, C.; Agoulon, A.; Barker, S.C.; Heylen, D.; Toty, C.; McCoy, K.D.; Plantard, O.; Rispe, C. A transcriptome-based phylogenetic study of hard ticks (Ixodidae). *Sci. Rep.* **2019**, *9*, doi:10.1038/s41598-019-49641-9.
16. Kobayashi, D.; Murota, K.; Itokawa, K.; Ejiri, H.; Amoa-Bosompem, M.; Faizah, A.N.; Watanabe, M.; Maekawa, Y.; Hayashi, T.; Noda, S.; et al. RNA virome analysis of questing ticks from Hokuriku District, Japan, and the evolutionary dynamics of tick-borne phleboviruses. *Ticks Tick-Borne Dis.* **2020**, *11*, 101364, doi:10.1016/j.ttbdis.2019.101364.

17. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549, doi:10.1093/molbev/msy096.
18. Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552, doi:10.1093/oxfordjournals.molbev.a026334.
19. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274, doi:10.1093/molbev/msu300.
20. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermini, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589, doi:10.1038/nmeth.4285.
21. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522, doi:10.1093/molbev/msx281.
22. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259, doi:10.1093/nar/gkz239.
23. Conow, C.; Fielder, D.; Ovadia, Y.; Libeskind-Hadas, R. Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms Mol. Biol.* **2010**, *5*, 16, doi:10.1186/1748-7188-5-16.
24. Tamura, K.; Nei, M.; Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci.* **2004**, *101*, 11030–11035, doi:10.1073/pnas.0404206101.
25. Zuckerkandl, E.; Pauling, L. Molecules as documents of evolutionary history. *J. Theor. Biol.* **1965**, *8*, 357–366, doi:10.1016/0022-5193(65)90083-4.
26. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **1997**, *13*, 555–556, doi:10.1093/bioinformatics/13.5.555.
27. Díaz-Sánchez, S.; Estrada-Peña, A.; Cabezas-Cruz, A.; de la Fuente, J. Evolutionary Insights into the Tick Hologenome. *Trends Parasitol.* **2019**, *35*, 725–737, doi:10.1016/j.pt.2019.06.014.
28. Blomström, A.-L.; Luz, H.R.; Öhlund, P.; Lukenge, M.; Brandão, P.E.; Labruna, M.B.; Berg, M. Novel Viruses Found in Antricola Ticks Collected in Bat Caves in the Western Amazonia of Brazil. *Viruses* **2019**, *12*, doi:10.3390/v12010048.
29. Mans, B.J.; Featherston, J.; Kvas, M.; Pillay, K.-A.; de Klerk, D.G.; Pienaar, R.; de Castro, M.H.; Schwan, T.G.; Lopez, J.E.; Teel, P.; et al. Argasid and ixodid systematics: Implications for soft tick evolution and systematics, with a new argasid species list. *Ticks Tick-Borne Dis.* **2019**, *10*, 219–240, doi:10.1016/j.ttbdis.2018.09.010.
30. Josek, T.; Walden, K.K.O.; Allan, B.F.; Alleyne, M.; Robertson, H.M. A foreleg transcriptome for Ixodes scapularis ticks: Candidates for chemoreceptors and binding proteins that might be expressed in the sensory Haller's organ. *Ticks Tick-Borne Dis.* **2018**, *9*, 1317–1327, doi:10.1016/j.ttbdis.2018.05.013.
31. Moutailler, S.; Popovici, I.; Devillers, E.; Vayssier-Taussat, M.; Eloit, M. Diversity of viruses in Ixodes ricinus, and characterization of a neurotropic strain of Eyach virus. *New Microbes New Infect.* **2016**, *11*, 71–81, doi:10.1016/j.nmni.2016.02.012.
32. Vandegrift, K.J.; Kapoor, A. The Ecology of New Constituents of the Tick Virome and Their Relevance to Public Health. *Viruses* **2019**, *11*, doi:10.3390/v11060529.