Mosquito community composition shapes virus prevalence patterns along anthropogenic disturbance gradients

Kyra Hermanns^a, Marco Marklewitz^a, Florian Zirkel^{b*}, Anne Kopp^a, Stephanie Kramer-Schadt^{c,d}, Sandra Junglen^{a,#}

^aInstitute of Virology, Charité – Universitätsmedizin Berlin, corporate member of Free University Berlin, Humboldt-University Berlin, and Berlin Institute of Health, Berlin, Germany;

^bInstitute of Virology, University of Bonn Medical Center, Bonn, Germany;

^cDepartment of Ecological Dynamics, Leibniz Institute for Zoo and Wildlife Research (IZW), Berlin, Germany;

^dInstitute of Ecology, Technische Universität Berlin, Berlin, Germany

Running title: Mosquito community composition shapes virus prevalence

* Present address: Florian Zirkel, Biotest AG, Dreieich, Germany

Correspondence to Sandra Junglen, sandra.junglen@charite.de

Abstract:

Previously unknown pathogens often emerge from primary ecosystems, but there is little knowledge on the mechanism behind. Most studies analyzing the influence of land-use change on pathogen emergence focus on a single host-pathogen system and often observe contrary effects. We studied virus diversity and prevalence patterns in natural and disturbed ecosystems using a multi-host and multi-taxa approach.

We detected 331 viral sequences pertaining to 49 viruses of ten RNA-virus families. Highest host and virus diversity was observed in pristine and intermediately disturbed habitats. The majority of the viruses was detected with low prevalence. However, nine viruses were found frequently of which five viruses increased in prevalence from pristine to disturbed habitats, in congruence with the dilution effect hypothesis. Interestingly, the observed increased prevalence of these five viruses in disturbed habitats was not caused by higher host infection rates but by increased host abundance, an effect tentatively named abundance effect.

Impact statement

Our data show that ecosystem disturbance can lead to a turnover in host community composition and that more individuals of a single species is a key driver of virus emergence.

Introduction:

A major challenge in understanding the emergence of infectious disease is to identify the driving factors responsible for changes in infectious disease dynamics. New infectious diseases mostly emerge in tropical regions which have undergone strong ecological and economic change (1). Tropical rainforests are terrestrial ecosystems with a high biodiversity, constituting a habitat with a diverse range of hosts, vectors and pathogens. This high host richness likely corresponds to a high pathogen richness, as each host is likely to carry its own specific pathogens (2). Pristine rainforests are subject to large scale anthropogenic land use transformation leading to increased contact between humans, wildlife, and pathogens endemic in the wildlife population (3, 4). Disturbed habitats often show a drastic decline in biodiversity or turnover of host species community composition, which is accompanied by an increase of species that are resilient to disturbance, so called generalist species. How these changes in community composition and pathogen abundance influence infection risks of humans and livestock are still unclear and a matter of scientific debate (5).

In this context, the dilution effect hypothesis postulates that a high diversity of different host species dilutes the prevalence of a specific pathogen in intact ecosystems, as the density of competent hosts (hosts that contribute to pathogen transmission and maintenance) is diluted by the presence of non-competent hosts. Biodiversity loss increases the infection prevalence of this specific pathogen in the disturbance-resilient, competent host species (6). Prerequisites for the occurrence of a dilution effect are the presence of species which differ in their susceptibility for a particular pathogen and a lower risk of extinction of competent hosts under habitat disturbance. This may then lead to a higher relative frequency of competent host species in modified habitats (7-9). Hence, in the case of zoonotic pathogens, the disease risk for hu-

mans is increased (6, 10). Studies confirming the dilution effect hypothesis focused on pathogens known to cause outbreaks in humans and livestock (e. g. West Nile virus (WNV), Sin Nombre virus and *Borrelia* spp.) and use generalist species as host (11-14). However, it seems that this effect cannot be generalized and that diverse mechanisms regulate infectious disease transmission dynamics in a host-, pathogen- and situation-dependent manner (15, 16). Some studies observed a contrary effect for WNV, Usutu virus and *Borellia* spp., where the infection rate or density of infected hosts increased with biodiversity, referred to as amplification effect (17-19).

These conflicting results indicate that biodiversity loss has complex impacts on disease risk and effects can be heterogeneous and scale-dependent (16, 20-22). Partial views on single host species or single pathogens cannot reveal general mechanisms. Existing studies either focused only on pathogen discovery aside from an ecological context or on one specific pathogen in a specific host across different habitats (12, 13, 17, 23). Studies with a wider scope analyzing community composition of entire host groups and assessing the genetic diversity of their pathogens in undisturbed and disturbed habitats are lacking and would be paramount for a comprehensive understanding of biodiversity-infection relationships. To this end, analyzing species interactions is crucial for a mechanistic understanding of the factors governing emerging infectious diseases (24).

Here, we provide a comprehensive analysis combining fields of community ecology and virology to understand how viral richness depends on host richness, and which role host-habitat associations play in viral prevalence patterns. Several arthropod-borne viruses (arboviruses) are prominent examples for emerging viruses that originated in biodiverse habitats and started from there their spread to new territories where they caused outbreaks and epidemics, e.g. dengue virus (DENV), yellow fever virus (YFV) and Zika virus (ZIKV) (25, 26). Nearly all

arboviruses are RNA viruses, which can rapidly evolve facilitating the acquisition of new hosts and geographic spread (27). Mosquitoes can transmit arboviruses during blood-feeding to vertebrates and thereby act as virus vectors. In addition, they can also be infected with their own specific viruses, which cannot be transmitted to vertebrates and are called insect-specific viruses (28). These may serve as models to study effects of land use change and changes in host community composition on mosquito-borne viruses.

In a preliminary analysis we had demonstrated that the prevalence of three insect-specific viruses isolated from mosquitoes sampled along an anthropogenic disturbance gradient in Côte d'Ivoire, West Africa, increased from pristine to disturbed habitat types (29). Subsequently, a plenitude of previously unknown RNA viruses was identified in these samples establishing novel species, genera and even families (30-41). To this end, we established broad-range generic PCR-assays for all viruses identified previously in these samples, as well as for all major virus taxa containing arthropod-associated viruses. With this, we aimed to assess the viral genetic diversity along with the prevalence patterns of each detected virus in an entire family of hosts (*Culicidae*) sampled across habitat types. We expect a higher species diversity in mosquitoes of the family *Culicidae* in habitats of low or no disturbance and predict a turnover in mosquito species composition along the disturbance gradient concomitant with an increase in disturbance-resilient mosquito species and their viruses. We thus hypothesize that habitat perturbation and subsequent changes in community composition influence virus transmission dynamics. This multi-host and multi-taxa study provides insight into common and distinct micro-evolutionary patterns of virus emergence and geographic spread at the interface of pristine and modified landscapes.

Materials and Methods:

Mosquito collection. In total, 4562 female mosquitoes were collected in five habitat types along an anthropogenic disturbance gradient in the Taï National Park in Côte d'Ivoire in 2004 (29). Habitat types represented pristine forest (PF), secondary forests (SF), agricultural areas (A), camps within primary forest (C) and villages (V) in the order of low to high human disturbance. Mosquitoes were identified morphologically (29) and based on their COI sequences (42). Mosquito heads were homogenized in 430 pools consisting of 1 to 50 individuals according to species and sampling location (29) resulting in 98 pools (764 mosquitoes) from the primary forest, 98 pools (1083 mosquitoes) from the secondary forest, 100 pools (1153 mosquitoes) from agricultural areas, 66 pools (568 mosquitoes) from research camps as well as 68 pools (994 mosquitoes) from two villages.

RT-PCR screenings and sequencing. RNA was extracted from the pooled supernatants using the QIAamp Viral RNA Mini Kit (Qiagen). The SuperScript™ III Reverse Transcriptase (Invitrogen - Thermo Fisher Scientific, Waltham, USA) was used for cDNA synthesis according to the manufacturer's instructions. Generic RT-PCR assays were established based on alignments of the RNA-dependent RNA polymerase (RdRp) sequences for the following taxa, peribunyaviruses, junciviruses, feraviruses, rhabdoviruses, flaviviruses, iflaviruses, orbiviruses and Cimodo virus. The RT-PCRs were designed based on sequence information from viruses, which have been previously isolated from these mosquitoes in cell culture (31-40), and from all major arbovirus taxa. Primer sequences and cycling conditions are available upon request. In addition, we used previously described assays for mesoniviruses (40), phenuiviruses (43), flaviviruses (44, 45) and alphaviruses (30) to test the samples. This entire approach allowed the detection of viruses regardless of isolation success in cell culture. PCR products were sequenced by Sanger sequencing (Microsynth AG, Balgach, Switzerland). In addition to the generic assays, specific

qPCRs were used to test for Cavally virus, Nsé virus and Mikado virus (40, and see below). The entire RdRp motifs of the third conserved region were amplified from all sequences with more than 5% divergence to another sequence using primer walking. Selected virus isolates were sequenced by metagenomic sequencing as previously described (30).

Genomic and phylogenetic analyses. All sequences were assembled and analyzed in Geneious R9.1.8 (46). Viral sequences were categorized based on genetic similarity and phylogenetic analysis. Sequences were compared to the NCBI database using blastn and blastx. Sequences with less than 95% amino acid identity to known viruses were considered as putative novel virus species and named Cimo virus (acronym for Côte d'Ivoire and mosquito) with ascending numbering. Viruses isolated in cell culture and those with completely sequenced genomes received individual names. For phylogenetic analyses, the amino acid sequences (families *Phenuiviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Rhabdoviridae* and *Iflaviridae*) or the nucleotide sequences (genera *Orbivirus*, *Flavivirus* and *Alphavirus* and family *Mesoniviridae*) of the detected viruses and the related established virus species were aligned by MAFFT-E v7.308 (47) in Geneious. An optimized maximum-likelihood phylogenetic tree with the substitution model based on Smart Model Selection (48) as implemented in PhyML (mainly LG or GTR, respectively) and 1,000 bootstrap replicates was calculated using PhyML (49). For the detected jonviruses the nonsynonymous and synonymous substitution rates were inferred using FEL (Fixed Effects Likelihood) as implemented in Datamonkey (50).

Virus growth analyses. Viruses were isolated in cell culture as described before (29). For the four novel viruses, Sefomo virus, Mikado virus, Tafomo virus and Sassandra virus, growth kinetics were performed using the mosquito cell line C6/36 (ECACC 89051705, *Aedes albopictus*). C6/36 cells were cultivated and seeded as previously described (51). The cells were infected at a MOI of 0.1 and incubated for 3 days at 28, 30, 32 and 34°C. Every 24 hours 75 µl

supernatant was taken and RNA was extracted using the NucleoSpin RNA Virus kit (Macherey-Nagel). cDNA was synthesized using the SuperScript IV reverse transcriptase (Invitrogen – Thermo Fisher Scientific) according to the manufacturer’s instructions. Specific quantitative RT-PCRs were established for all four viruses. Respective primer and probe sequences are

SefomoV-F: 5’-TGGTTGAGACCTTCTGAGACTTTTC-3’, SefomoV-R: 5’-CAAAGGCCATCCCGAAGTATC-3’, SefomoV-TM: 5’-6-FAM/CCATTACAC/ZEN/CTCATCCCTATTTTCATGCTGG/Iowa Black®-FQ-3’, MikadoV-F: 5’-GA-GAACTGTCAAAAATGGAGAAGAGA-3’, MikadoV-R: 5’-AGATGGCACCATTTCAGTGATATAG-3’, MikadoV-TM: 5’-6-FAM/GCCAACAGC/ZEN/CAATTAAGAGAATGA/Iowa Black®-FQ-3’, TafomoV-F: 5’-AATCTGATCTGGAGGACGAGTTG-3’, TafomoV-R: 5’-GCTGTTGATTAGCTGTG-CATGAT-3’, TafomoV-TM: 5’-6-FAM/GGTTCTTGC/ZEN/TGGACCAGGTG/Iowa Black®-FQ-3’, SassandraV-F: 5’-CATTTTGGGAAGGAGATTTTTCGA-3’, SassandraV-R: 5’-GATCAAATTTCCAA-TAGCCCATAAA-3’ and SassandraV-TM: 5’-6-FAM/TGGACCTCA/ZEN/AGCGGATTCAAC-CGT/Iowa Black®-FQ-3’.

Statistical analysis. *Virus prevalence.* To estimate virus prevalence, the minimum infection rate (MIR) and the maximum likelihood estimation (MLE) per 1000 mosquitoes was calculated with the Excel Add-In PooledInfRate, version 4.0 (52). Virus data analyses were performed in GraphPad Prism 7.04 (GraphPad Software, San Diego, USA).

Biodiversity analyses. All biodiversity analyses were done with R version 3.6.2 (53). We used rarefaction curves and Hill numbers of diversity order $q = 0$ (species richness) to compare mosquito communities between the five habitats ((54); R-package ‘iNEXT’: (55, 56)). To this end, we summed the number of mosquito species per habitat and excluded non-defined species from the analysis. For detection of underlying gradients, we used the Bray-Curtis dissimilarity index ((57); R-package ‘vegan’ (58)) and hierarchical cluster analysis to group main mosquito

species according to their habitat association (R-package ‘pheatmap’). To assess whether there are significant associations between these mosquito groups and habitat, we fitted a log-linear model (generalized linear model with ‘log’ link and Poisson error distribution) to the number of mosquitoes counted per group as response variable, with mosquito group and habitat as covariates in interaction. Since there were no counts of *Coquillettidia* species in camp and village, we added a small constant of 1 to all responses to avoid singularities in the model outcome and to be able to model it with a log transformation. We assessed significance of the single terms using a log likelihood ratio test (LRT) of this saturated model.

We used Spearman’s rho on the full dataset to assess which viruses were associated with each other in the different pools. We focused our analyses on abundant viruses we had found at least 10 times in all pools. For assessing virus-habitat relationships expressed as the probability of detecting one of the abundant viruses in a specific habitat, we fitted a generalized linear mixed model with binomial error structure and logit link and the respective virus type detected (1) or not (0) as response. As covariates, we included the number of mosquitoes per pool and the habitat. We included the number of mosquitoes per pool as controlling variable, because the probability to detect any virus in a pool logically increases when more potential hosts are combined into a pool. Pairwise differences in virus infection probabilities in the habitats were assessed post hoc with Tukey contrasts (R-package multcomp).

Nucleotide Sequence Accession Numbers. The sequence fragments and viral genomes were assigned the GenBank accession numbers ... (*pending*), respectively. Accession numbers of representative sequences are listed in Table 1.

Results:

Biodiversity analyses. The dataset included 42 unique mosquito species units. We counted and rarefied (r given as asymptotic diversity estimate) the highest number of species (n) in the primary forest (PF, n = 20 | r = 38 | with i number of individuals = 462), followed by camp (C, n = 18 | r = 24 | i = 418) and secondary forest (SF, n = 14 | r = 24 | i = 651), agriculture (A, n = 17 | r = 18 | i = 882), and village (V, n = 13 | r = 15 | i = 857) (**Fig. 1**).

Figure 1: Mosquito species richness per habitat. Rarefaction curves (with 95% confidence intervals) show the observed (solid lines) and interpolated (dotted line) mosquito diversity in the five habitats.

Rarefaction curves were still increasing for camp, primary and secondary forest, indicating that a higher trapping effort would have led to a higher number of species. However, confidence intervals were strongly overlapping, especially for camp and primary forest and for agriculture, secondary forest and village. Bray-Curtis dissimilarity (BCd) yielded the highest difference in community assembly between primary forest and village (BCd = 0.90), and highest similarity between agriculture and secondary forest (BCd = 0.69). The hierarchical cluster analysis confirmed the finding that agriculture and secondary forest clustered together via *Culex decens* and *Culex nebulosus*, while villages and camp mainly clustered via *Culex nebulosus*. Agriculture had high amounts of *Culex annulioris* and *Coquillettidia metallica*, while primary forest mosquito communities were characterized by *Uranotaenia* species. Villages were mainly characterized by *Anopheles* species (**SI Fig. 1 a**). We therefore grouped our mosquitoes to the eight major groups *Culex decens* (i = 1101), *Culex nebulosus* (i = 719), *Culex annulioris* (i = 263), *Culex sp.* (i = 222), *Coquillettidia sp.* (i = 95), *Uranotaenia sp.* (i = 502) and *Anopheles sp.* (i = 294), while all others, mainly the non-defined ones, were termed ‘others’ (i = 1366). Based on these findings, our anthropogenic disturbance gradient was as follows (from low to high disturbance): PF-SF-A-C-V. Significant associations were found between main mosquito groups and habitat (LRT $p < 0.001$, df = 28, deviance = 3782; **SI Table 1 and SI Fig. 1 b**).

Assessment of the genetic virus diversity. In total, we found 331 viral RdRp sequences pertaining to 34 putative novel viruses and to 15 previously identified viruses of the families *Phenuiviridae*, *Peribunyaviridae* and *Phasmaviridae* of the order *Bunyavirales*, as well as of the families *Flaviviridae*, *Togaviridae*, *Rhabdoviridae*, *Reoviridae*, *Iflaviridae* and *Mesoniviridae* (**Table 1**). Sequences with at least 5% pairwise amino acid distance to known viral RdRp sequences were suggested to pertain to distinct viral species.

Table 1: Distribution and host-association of detected viruses. Number of positive pools per habitat and mosquito host species of all detected viruses and virus-like sequences. The main mosquito host species are indicated in bold letters.

	Virus	No. of positive Pools						Mosquito species	Isolate	Representative pool (GenBank accession number) [†]	sequence length [nt]	References
		Σ	PF	SF	A	C	V					
Bunyviridae	Gouléako virus	33	2	4	5	6	16	<i>Culex nebulosus</i> , <i>Culex decens</i> , <i>Culex</i> spp., nd*	x	A05 (NC_043051)	full genome	[35]
	Cimo phenivir I	1	0	0	0	1	0	nd		A27		
	Cimo phenivir II	12	8	1	0	3	0	<i>Uranotaenia mashaensis</i> , <i>Uranotaenia ornata</i> , <i>Uranotaenia</i> spp., nd		B02		
	Cimo phenivir III	2	2	0	0	0	0	nd		B14		
	Cimo phenivir IV	1	1	0	0	0	0	nd		B98		
	Cimo phenivir V	1	0	0	1	0	0	<i>Anopheles</i> spp.		E02		
	Cimo phenivir VI	6	0	0	0	0	6	<i>Anopheles gambiae</i> , <i>Anopheles nill</i> , <i>Anopheles</i> spp., nd		F09		
	Sefomo virus	1	0	1	0	0	0	<i>Culex decens</i>	x	C43	full genome	
	Phasi Charoen-like virus	2	0	0	1	0	1	<i>Aedes aegypti</i> , nd		F04		
								<i>Culex nebulosus</i> , <i>Culex decens</i> , <i>Culex</i> spp.,				
	Herbert virus	42	3	8	8	6	17	<i>Mimomyia mimomyiaformis</i> , <i>Coquillettia</i> spp., nd	x	F23 (NC_038714)	full genome	[37]
	Tai virus	3	0	2	1	0	0	<i>Culex (Culex) decens</i> , nd	x	F47 (NC_034459)	full genome	[37]
	Cimo peribunyavirus I	4	3	0	1	0	0	<i>Uranotaenia mashaensis</i> , nd		B04		
	Cimo peribunyavirus II	1	0	0	1	0	0	nd		D55	3596 (L) 2228 (M)	
	Cimo peribunyavirus III	2	0	1	0	1	0	<i>Culex nebulosus</i> , <i>Culex decens</i>		A07		
	Ferak virus	20	1	3	4	6	6	<i>Culex nebulosus</i> , <i>Culex decens</i> , nd	x	C51 (NC_043031)	full genome	[36]
	Jonchet virus	17	2	0	0	9	6	<i>Culex decens</i> , <i>Culex nebulosus</i> , <i>Culex</i> spp., nd	x	B81 (NC_038706)	full genome	[36]
	Spilkins virus	12	1	2	2	2	5	<i>Culex nebulosus</i> , <i>Culex</i> spp., nd		A28	complete CDS	
	Mikado virus	3	0	0	3	0	0	<i>Culex annulirois</i>	x	D35	complete CDS	
Rhabdoviridae	Cimo rhabdovirus I	40	0	20	15	3	2	<i>Culex decens</i> , <i>Culex</i> spp., nd		A02		
	Cimo rhabdovirus II	1	0	0	0	1	0	nd		A30		
	Cimo rhabdovirus III	2	2	0	0	0	0	nd		B58		
	Cimo rhabdovirus IV	2	0	1	1	0	0	nd		C68		
	Cimo rhabdovirus V	3	0	0	3	0	0	<i>Coquillettia metallica</i> , nd		D24		
potential NIRVS	Rhabdovirus-like NIRVS I	1	0	0	0	1	0	nd		A19		
	Rhabdovirus-like NIRVS II	1	1	0	0	0	0	nd		B82		
Reoviridae	Cimodo virus	5	0	5	0	0	0	<i>Culex decens</i> , nd	x	C74 (NC_023420)	full genome	[31]
	Wanken orbivirus	5	5	0	0	0	0	<i>Uranotaenia mashaensis</i> , nd		B14		
Mesoniviridae	Cavally virus	30	3	5	4	3	15	<i>Culex nebulosus</i> , <i>Culex decens</i> , <i>Culex</i> spp., nd	x	C79 (NC_015668)	full genome	[39]
	Hana virus	1	0	0	0	1	0	<i>Culex</i> spp.	x	A04 (NC_020899)	full genome	[40]
	Méno virus	1	0	0	1	0	0	<i>Uranotaenia</i> spp.	x	E09 (NC_020900)	full genome	[40]
Togaviridae	Nsé virus	7	1	3	1	1	1	<i>Culex nebulosus</i> , <i>Culex decens</i>	x	F24 (NC_020901)	full genome	[40]
	Tai Forest alphavirus	1	0	1	0	0	0	<i>Culex decens</i>		C21 (NC_032681)	full genome	[30]
Iflaviridae	Cimo iflavirus I	2	0	1	1	0	0	<i>Culex decens</i> , <i>Culex nebulosus</i>		D52		
	Cimo iflavirus II	1	0	1	0	0	0	<i>Culex decens</i>		C61		
	Cimo iflavirus III	2	0	2	0	0	0	<i>Culex decens</i>		C95		
	Sassandra virus	2	0	1	0	0	1	<i>Culex</i> spp., nd	x	C93	full genome	
	Niénokoué virus	9	2	3	3	0	1	<i>Coquillettia metallica</i> , <i>Culex</i> spp., nd	x	B51 (NC_024299)	full genome	[33]
Flaviviridae	Nounané virus	3	3	0	0	0	0	<i>Uranotaenia mashaensis</i>	x	B31 (NC_033715)	complete CDS	[32]
	Anopheles flavivirus	3	0	0	0	1	2	<i>Anopheles gambiae</i> , <i>Anopheles</i> spp.		F10		
	Tafomo virus	7	2	1	2	2	0	<i>Culex</i> spp., nd	x	B22	full genome	
	Cimo flavivirus I	13	4	6	3	0	0	<i>Coquillettia</i> spp. (Unknown COI-type C69), nd		B36		
	Cimo flavivirus II	4	1	0	3	0	0	<i>Uranotaenia</i> spp., nd		D01		
	Cimo flavivirus III	5	2	0	0	2	1	<i>Uranotaenia mashaensis</i> , <i>Uranotaenia</i> spp., nd		B01		
	Cimo flavivirus IV	5	0	0	4	0	1	<i>Mimomyia</i> spp., nd		E08		
	Cimo flavivirus V	3	0	0	2	0	1	<i>Mimomyia hispida</i> , nd		D20		
	Cimo flavivirus VI	3	0	0	2	0	1	<i>Anopheles rhodesiensis rupicolus</i> , <i>Anopheles</i> spp.		E02		
	Cimo flavivirus VII	3	1	2	0	0	0	nd		B36		
	Cimo flavivirus VIII	2	1	1	0	0	0	<i>Eretmapodites intermedius</i> , nd		B85		
	Cimo flavivirus IX	1	0	1	0	0	0	nd		C51		
	Cimo flavivirus X	1	0	0	1	0	0	nd		E08		
	Cimo flavivirus XI	1	0	0	0	0	1	<i>Culex nebulosus</i>		F41		
potential NIRVS	Flavivirus-like NIRVS I	10	2	3	4	0	1	<i>Coquillettia metallica</i> , nd		B60		
	Flavivirus-like NIRVS II	2	0	0	1	0	1	<i>Aedes aegypti</i> , <i>Aedes</i> spp.		E01		
	Flavivirus-like NIRVS III	3	0	2	0	1	0	<i>Eretmapodites</i> spp.		C78		
	Flavivirus-like NIRVS IV	2	2	0	0	0	0	nd		B18		
	Flavivirus-like NIRVS V	2	1	0	0	1	0	nd		B68		
	Flavivirus-like NIRVS VI	2	0	0	0	0	2	<i>Mimomyia</i> spp., nd		F16		
	Flavivirus-like NIRVS VII	3	0	2	1	0	0	nd		D57		
	Flavivirus-like NIRVS VIII	13	8	1	2	2	0	<i>Uranotaenia</i> spp., nd		A25		
	Flavivirus-like NIRVS IX	8	2	1	2	2	1	<i>Uranotaenia</i> spp., <i>Uranotaenia mashaensis</i> , nd		E26		
	Flavivirus-like NIRVS X	15	4	3	6	1	1	<i>Coquillettia metallica</i> , nd		E11		

*nd -> not determined, [†]RdRp encoding segment (previously published sequences are indicated in italic)

The family *Phenuiviridae* (order *Bunyavirales*) includes important arboviruses within the genus *Phlebovirus* but also numerous insect-specific viruses that for example belong to the genera *Goukovirus* and *Phasivirus* (59). Nine distinct phenuiviruses were detected, which included seven novel viruses, named Sefomo virus (acronym for secondary forest mosquito virus) and Cimo phenuivirus I-VI, as well as Gouléako virus (GOLV) and Phasi Charoen-like virus (PCLV) (35, 60). For all viruses, the number of positive samples per habitat is summarized in **Table 1**. Phylogenetic analyses showed that the viruses grouped with insect-specific viruses of the genera *Goukovirus*, *Phasivirus*, *Hudivirus* and *Beidivirus*, as well as with the unclassified insect viruses related to the uncultured virus isolate acc 9.4 (**Fig. 2a**). Interestingly, Cimo phenuivirus V branched basal to tenuiviruses that are transmitted between plants by planthoppers (61). Cimo phenuivirus V was found in *Anopheles* spp. mosquitoes collected at a rice plantation. At this point we cannot differentiate whether the mosquito ingested infected plant material or if this novel tenui-like virus can infect mosquitoes.

The family *Peribunyaviridae* (order *Bunyavirales*) consists of two arbovirus genera (*Orthobunyavirus* and *Pacuvirus*) and two genera which contain insect-restricted viruses (*Herbevirus* and *Shangavirus*) (62). We identified two novel peribunyaviruses, named Cimo peribunyavirus I and II. These viruses formed a monophyletic clade that shared a most recent common ancestor with arboviruses of the genera *Orthobunyavirus* and *Pacuvirus* (**Fig. 2b**). In addition, the insect-specific herbeviruses Taï virus and Herbert virus (HEBV) (37), as well as a previously undescribed herbevirus, named Cimo peribunyavirus III, were detected that fell into the clade of herbeviruses (**Fig. 2b**).

The family *Phasmaviridae* (order *Bunyavirales*) comprises only insect-specific viruses (59). We found two prototype species of the genera *Feravirus* and *Jonvirus*, Ferak virus (FERV) and Jonchet virus (JONV), which were previously isolated in cell culture from these mosquitoes

(36). While a great diversity of novel phasmaviruses has been found since the first discovery of this family, no additional members of the genus *Jonvirus* have been identified (59). Here, we further detected two previously unknown jonviruses, named Mikado virus and Spilikins virus, which are closely related to the prototype virus JONV in phylogenetic analyses (**Fig. 2c**). For both novel jonviruses, the complete coding sequence (CDS) was sequenced and phylogenies of the M and S segments are shown in the **Supplementary Information (SI) Figure 2**. In all phylogenies, the two novel viruses cluster well supported with JONV.

Figure 2: Phylogenetic analyses of detected bunyaviruses. Phylogenetic trees were inferred with PhyML (LG substitution model) based on MAFFT-E protein alignments covering the conserved RdRp motifs of the families *Phenuiviridae* (A), *Peribunyaviridae* (B) and *Phasmaviridae* (C). Novel viruses from this study are indicated in red and previously published viruses detected in our data set are indicated in blue. Sample origin from the different habitat types is indicated by colored circles while no detection is indicated by grey circles. Live virus isolates are marked with a blue asterisk. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

The family *Rhabdoviridae* (order *Mononegavirales*) is highly diversified and currently contains 20 genera. Rhabdoviruses infect vertebrates, arthropods and plants (63). We detected five previously unknown rhabdoviruses, named Cimo rhabdovirus I-V that clustered with different unclassified clades of mosquito-associated rhabdoviruses across the rhabdovirus phylogeny (**Fig. 3a**).

The family *Iflaviridae* (order *Picornavirales*) consists of a single genus whose members are restricted to arthropod hosts (64). We detected four novel iflaviruses, named Sassandra virus and Cimo iflavirus I-III, which were placed in three different clades in phylogenetic analyses (**Fig. 3b**). Cimo iflavirus I and III grouped with *Bombyx mori* iflavirus and other lepidopteran iflaviruses, while the short sequence fragment of Cimo iflavirus II did not form a well-supported clade with known iflaviruses. Sassandra virus clustered with two previously described iflaviruses from mosquitoes (**Fig. 3b**).

Figure 3: Phylogenetic analyses of detected rhabdoviruses and iflaviruses. Phylogenetic trees were inferred with PhyML (LG substitution model) based on MAFFT-E protein alignments covering the conserved RdRp motifs of the families *Rhabdoviridae* (A) and *Ifelviridae* (B). Novel viruses from this study are indicated in red and detected virus-like sequences are indicated in grey. Sample origin from the different habitat types is indicated by colored circles while no detection is indicated by grey circles. Live virus isolates are marked with a blue virion. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

The genus *Flavivirus* (family *Flaviviridae*) includes important arboviruses as well as viruses with a single host tropism for arthropods or vertebrates (65). Insect-specific flaviviruses can be divided into two groups. Classical insect-specific flaviviruses form a monophyletic clade in basal phylogenetic relationship to all other flaviviruses while dual-host affiliated insect-specific flaviviruses are phylogenetically affiliated with the arboviruses of this genus (28). We found twelve undescribed flaviviruses, named Tafomo virus (acronym for Tai forest mosquito virus) and Cimo flavivirus I-XI, as well as a strain of *Anopheles* flavivirus (66). All sequences clustered within the clade comprising the classical insect-specific flaviviruses in phylogenetic analysis (**Fig. 4a**). In addition, two previously characterized flaviviruses, the dual-host affiliated insect-specific flavivirus Nounané virus and the classical insect-specific flavivirus Niénokoué virus (NIEV), were detected in the mosquitoes (32, 33).

The family *Reoviridae* comprises 15 genera with highly variable biological properties. Three genera contain arboviruses (*Orbivirus*, *Coltivirus* and *Seadornavirus*) while viruses belonging to other genera infect vertebrates, plants, fungi or insects (67). One novel orbivirus, named Wanken orbivirus (WKOV), was detected in the mosquitoes. The complete genome of WKOV was sequenced and the polymerase encoded on the first segment showed 56% pairwise amino acid identity to Ninarumi virus. Phylogenetic analysis of the polymerase (VP1) placed WKOV on a branch with Ninarumi virus, a virus detected in *Aedes fulvus* mosquitoes collected in Peru (68). This clade was placed basal to the clade of *Culicoides*- and *Phlebotominae*-transmitted orbiviruses (e.g. Bluetongue virus) and Letea virus, a virus detected in snakes (*Natrix natrix*) captured in Romania (69) (**Fig. 4b**). Additional phylogenetic analyses based on VP3, VP4, VP5 and VP7 also placed WKOV together with Ninarumi virus and Letea virus basal to the clade of *Culicoides*- and *Phlebotominae*-transmitted orbiviruses (**SI Fig. 3**). Additionally, Cimodo virus,

333 which most likely defines a novel reovirus genus and which was previously characterized (31),
334 was found (**SI Fig. 4a**).
335

Figure 4: Phylogenetic analyses of detected flaviviruses and orbiviruses. Phylogenetic trees were inferred with PhyML (GTR substitution model) based on MAFFT-E nucleotide alignments covering the conserved RdRp motifs of the genera *Flavivirus* (A) and *Orbivirus* (B). Novel viruses from this study are indicated in red, previously published viruses detected in our data set are indicated in blue and detected virus-like sequences are indicated in grey. Sample origin from the different habitat types is indicated by coloured circles while no detection is indicated by grey circles. Live virus isolates are marked with a blue virion. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

Most viruses of the genus *Alphavirus* (family *Togaviridae*) are arboviruses. Contrary to the genus *Flavivirus*, only few insect-specific alphaviruses have been discovered thus far (70). We detected the previously characterized insect-restricted Taï Forest alphavirus (30) but no further alphaviruses were found (**SI Fig. 4b**).

The family *Mesoniviridae* (order *Nidovirales*) consists of insect-specific viruses (40, 71). The four previously described mesoniviruses Cavally virus (CAVV), Nsé virus, Hana virus, and Méno virus were detected in the mosquitoes (40) (**SI Fig. 4c**).

Besides the 12 previously published virus isolates from this sampling (31-41), Sefomo virus, Mikado virus, Sassandra virus, and Tafomo virus were isolated in C6/36 mosquito cells indicating the detection of functional viruses. The viruses replicated well in C6/36 cells incubated at 28-30°C but virus growth was impaired at 32°C for all but Mikado virus. No virus could replicate at 34°C indicating the detection of insect-restricted viruses (**Fig. 5**).

Figure 5: Temperature-dependent replication of novel virus isolates. C6/36 cells were infected with an MOI of 0.1 with Sefomo virus (A), Mikado virus (B), Sassandra virus (C) and Tafomo virus (D) in duplicates. Replication was measured for 3 dpi at 28, 30, 32 and 34°C.

Detection of non-retroviral integrated RNA virus sequences. Genome fragments of flavi-, bunya- and rhabdoviruses were found to have integrated into mosquito genomes and persist as so called non-retroviral integrated RNA virus sequences (NIRVS) (44, 72-74). We detected flavi- and rhabdovirus-like sequences with defective ORFs within the conserved region of the RdRp gene suggesting the detection of NIRVS. The two potential rhabdovirus-like NIRVS encoded either an internal stop codon (rhabdovirus-like NIRVS II – B82) or sequence elongation attempts resulted in rhabdovirus-like sequences that contained frame shifts (rhabdovirus-like NIRVS I – A19). We were further able to amplify these defective virus-like sequences directly

from nucleic acid extracts without prior cDNA synthesis suggesting that these sequences were present as DNA. No amplicates were obtained for the sequences with contiguous ORFs (SI Fig. 5a).

Furthermore, we obtained ten flavivirus-like sequences with frame shifts, internal stop codons, deletions or integrations (SI Fig. 5b). In case of flavivirus-like NIRVS II and III, parts of the potentially integrated sequences were related to sequences from insects including genome loci from *Aedes* mosquitoes. Flavivirus-like NIRVS II was amplified from an *Aedes aegypti* pool (E01) and the integrated fragment (68 nt) consisted of 56 nt with 81% identity to *Aedes aegypti* steroid hormone receptor homolog (AaHR3-2) gene and a 12 nt duplication of the sequence immediately adjacent to the integration site. In addition to the interrupted flavivirus-like sequence in pool E01, an identical continuous flavivirus sequence was obtained from the same pool. A similar observation was made with flavivirus-like NIRVS I (pool B60). The sequence of flavivirus-like NIRVS III (pool C78) was detected in *Eretmapodites intermedius* and undetermined mosquitoes. This sequence was profoundly defective (frameshifts, internal stop codon and deletions) and continued into a 416 nt long sequence with low similarity to *Aedes albopictus* and *Apis* spp. genome loci (SI Fig. 5b). The very few and short sequences of *Eretmapodites* spp. available in GenBank most likely impeded the identification of the host genome sequence of flavivirus-like NIRV III.

Detected viruses show a high host specificity. We next sought to identify if the detected viruses were associated with specific mosquito host species. In total, 43% of the pools were found positive for any virus. The gross majority of the viruses (n= 39) was detected in a single mosquito species and only 10 viruses were detected in two or three different closely related

mosquito species which belonged to the same mosquito genus indicating a high host specificity (**Table 1**). Unfortunately, ordination techniques like non-metric multidimensional scaling did not converge due to too few positive findings of viruses in the pools (data not shown). We often detected multiple viruses in one mosquito pool, especially in *Culex* pools, indicating possible co-infections mainly between GOLV, HEBV, CAVV and FERV. These viruses were all associated with *Culex nebulosus* as main mosquito host and consequently often found together in pooled mosquitoes of this species. We found significant positive associations between GOLV and HEBV, as well as between GOLV and CAVV (**Fig. 6**). The association between GOLV and HEBV strongly increased (Spearman's $\rho = 0.84$, $p < 0.05$) when considering only the host mosquito *Culex nebulosus*. Here, also correlations of both viruses with FERV and Spilkins virus increased ($\rho \sim 0.45$, $p < 0.01$; results not shown).

Figure 6: Spearman's rank correlation rho for the most abundant viruses. Only significant correlations are shown. AnthroDist refers to the gradient of anthropogenic disturbance from low to high: PF, primary forest; SF, secondary forest; A, agriculture; C, camp, and V, village.

Whether these mixed infections result from multiple infected individuals or from co-infected single mosquitoes could not be analyzed, as no homogenates of individual specimens were available. All four viruses could be isolated together in cell culture from multiple pools. The

replication of all four viruses in co-infected cell cultures was confirmed by quantitative real-time PCR suggesting no general inhibitory effect and the possibility of simultaneous infection of a single mosquito.

The three jonviruses, JONV, Spilikins virus and Mikado virus, were each also associated with specific mosquito species, namely with *Culex decens*, *Culex nebulosus* and *Culex annulioris*, respectively. Interestingly, their partial RdRp sequences showed a high degree of variation in the third codon positions which did not alter the translated protein sequences (called synonymous substitutions). For example, JONV and Mikado virus diverged by approximately 20% in their nucleotide (nt) sequences but only by 7% in their amino acid (aa) sequences. Similarly, Spilikins virus and JONV showed nt and aa divergences of 25% and 12%-15%, respectively. FEL analysis of 31 detected jonvirus sequences found significant (p-value <0.05) evidence of negative selection for 88 out of 117 codons. These findings could point towards adaptation to a specific mosquito host under purifying selection. Similar strains with mainly variation in the third codon position were also observed for two Cimodo virus strains, HEBV and Cimo peribunyavirus III, two strains of Cimo peribunyavirus I, two strains of Anopheles flavivirus, as well as Cimo flavivirus I and Cimo flavivirus VII.

Virus prevalence patterns along the anthropogenic disturbance gradient.

According to current hypotheses in infectious disease ecology, ecological perturbation and changes in community composition are expected to influence transmission dynamics of infectious diseases (3, 5, 13). We thus next analyzed the prevalence patterns of all detected viruses as well as abundance of their mosquito host species along the anthropogenic disturbance gradient. The abundance of different mosquito genera varied considerably across the different habitats, as described previously by Junglen et al. (29). The dominant genus across all habitats

except in the primary forest was *Culex* (50.5%), whereas in the primary forest mosquitoes of the genus *Uranotaenia* were most frequently sampled. *Culex decens* was most abundant in the secondary forest and agricultural areas, whereas *Culex nebulosus* mosquitoes were the most abundant species in villages and at camp sites located in the primary forest. A large fraction of 56.7%, 43.9% and 52.1% of the sampled mosquitoes in the primary and secondary forest as well as at the camp sites, respectively, could not be identified to species level either due to morphological damage or to limitations of taxonomic keys. The actual number of different species in these habitats is therefore likely to be higher and the relative abundance of some mosquito species may have been underestimated.

The highest virus richness was observed in the intermediately disturbed habitats secondary forest and agriculture (**Fig. 7a**). However, the proportional number of tested mosquitoes was lower in the primary forest so that the richness in this habitat is likely underestimated.

The majority of the viruses (82%, $n = 40$) occurred with a low frequency of less than ten positive samples in total (see **Table 1**). The cumulated MIR for all detected viruses was slightly higher in villages and at the camp sites compared to the other habitats (**Fig. 7b**). This effect was mainly caused by the increasing prevalence of several bunyaviruses and one mesonivirus (**SI Fig. 6a, b, f and h**), while other taxa like reoviruses and iflaviruses were found in specific habitats (**SI Fig. 6c and i**), or like flaviviruses and rhabdoviruses increased in prevalence towards pristine or intermediately disturbed habitats (**SI Fig. 6 e and g**). Of note, no amplification effect was observed in habitats with higher biodiversity (**Fig. 7a and b**).

Figure 7: Richness and cumulative MIR across all tested virus taxa. The number of distinct viruses (A) and the cumulated MIR per 1000 mosquitoes (B) were calculated for all habitat types and for the complete data set. Different virus taxa are shown in different colors.

However, nine viruses were found more frequently with detection rates ranging from 12 to 42 strains (Table 1). As the sample set consisted of pooled mosquito specimens, we estimated viral infection rates in the different habitats using the two approaches MIR and MLE which

showed similar results (**Fig. 8**). Four bunyaviruses (GOLV, HEBV, FERV and Spilikins virus) and one mesonivirus (CAVV), all mainly associated with *Culex nebulosus*, increased in prevalence towards disturbed habitat types (**Fig. 8a, b, c and e - left graphs and 8h**). A trend, that seemed to support the dilution effect hypothesis. In contrast, Cimo rhabdovirus I had its highest prevalence in the intermediately disturbed habitats (secondary forest and agricultural areas) (**Fig. 8d – left graph**) and two viruses increased in prevalence towards the primary and secondary forest (Cimo phenuivirus II and Cimo flavivirus I) (**Fig. 8f and i**). JONV prevalence slightly increased in the villages and more prominently at the camp sites compared to the other habitats (**Fig. 8g**). The prevalence patterns corresponded for eight of the nine viruses to the relative abundance of the main mosquito host species (**Fig. 8**). JONV was the only virus that showed no relationship between mosquito host abundance and virus prevalence suggesting that other factors may influence JONV abundance. Cimo rhabdovirus I and JONV were both mainly associated with *Culex decens* mosquitoes as host. However, in contrast to GOLV, HEBV, CAVV and FERV, which were frequently detected together in *Culex nebulosus* mosquitoes, Cimo rhabdovirus I and JONV were only twice found together in the same pool. This could hint to a possible interference between these viruses and might be a reason for the unusual prevalence pattern of JONV.

The probability of finding an infection with GOLV, HEBV and CAVV was significantly higher in villages ($p < 0.1$, **SI Table 2**). These viruses were mainly associated with *Culex nebulosus* mosquitoes. Cimo phenuivirus II was associated with *Uranotaenia* sp. mosquitoes in the primary forest ($p < 0.1$, **SI Table 2**), whereas JONV and FERV mainly occurred in the camps and villages, associated with *Culex nebulosus*, *Culex decens* and other mosquito species. Cimo flavivirus I did not show a clear tendency, but mainly occurred in the primary and secondary forest as well as the agricultural sites in *Coquillettidia* sp. and other mosquito species. Cimo rhabdovirus

495 I was almost exclusively associated with *Culex decens* and the detection probability signifi-
496 cantly increased in the secondary forest and the agricultural sites ($p < 0.1$, **SI Table 2**).
497

Figure 8: Prevalence patterns of selected viruses along the disturbance gradient. For all viruses, that were detected in >10 pools, GOLV (A), HEBV (B), FERV (C), Cimo rhabdovirus I (D), CAVV (E), Cimo phenivirus II (F), Jonchet virus (G), Spilikins virus (H) and Cimo flavivirus I (I),

the MIR and MLE per 1000 mosquitoes of the whole data set was calculated for all habitat types. The abundance of the main mosquito host species was plotted. The five viruses GOLV (A), HEBV (B), FERV (C), Cimo rhabdovirus I (D), and CAVV (E) occurred frequently enough in their main mosquito host species (>10 positive pools) to analyze their prevalence in these species. For these viruses, the MIR and MLE per 1000 mosquitoes of the respective species was calculated for all habitat types (right graphs). Significant differences in the infection probability with the most abundant viruses in the different habitats are shown in SI Table 2.

We next investigated infection rates only in the main mosquito host species across the different habitat types. Only five of the nine viruses were detected frequently enough in their main mosquito host species ($n > 10$) to calculate host-specific infection rates. Surprisingly, no trend of increasing or decreasing virus prevalence was detected along the disturbance gradient (**Fig. 8a-e – right graphs**) as would be expected in case of a dilution or amplification effect (13, 17). Viral infection rates did not change considerably in their mosquito hosts between disturbed and undisturbed habitat types. Notably, the increase in prevalence of specific viruses resulted from shifts in the mosquito community composition along the gradient, which caused increased abundance rates of the main mosquito host species and concomitantly higher prevalence rates of the viruses they were carrying. We thus refer to this observation as abundance effect (summarized in **Fig. 9**).

Figure 9: Schematic presentation of the abundance effect. The different habitat types along the anthropogenic disturbance gradient are depicted by photos and drawings. Infection rates are shown for GOLV in all sampled mosquitoes (representing MIR and MLE values as shown in Figure 8) and only in the main mosquito host species, *Culex nebulosus*. The abundance of the main mosquito host species, *Culex nebulosus* is indicated by a grey line.

Discussion:

In this study, we analyzed the interplay between host community composition, habitat disturbance and virus prevalence in a multi-host and multi-taxa approach. We characterized the genetic diversity of RNA viruses in an entire family of hosts (*Culicidae*), which were sampled along an anthropogenic disturbance gradient. We subsequently studied prevalence patterns

of all detected viruses per habitat type, as well as mosquito community composition per habitat type. We discovered an exceptionally high diversity of 49 distinct viruses, of which 34 were previously unknown members of seven different RNA virus families. We demonstrated that the majority of these viruses occurred at low minimum infection rates of 0.22 – 1.97 infected mosquitoes per 1000 tested mosquitoes. Nine viruses occurred more frequently across the disturbance gradient, of which five increased in prevalence from pristine to disturbed habitat types. We could show that the detection rates of these viruses corresponded to the abundance patterns of their specific mosquito host species. Importantly, the differences in virus prevalence were driven by the number of hosts present in a specific habitat type and not by changes in host infection rates. This interplay was named abundance effect. These data show that host community composition critically influences virus prevalence which has direct impact on our understanding of infectious disease emergence mechanisms.

The majority of the detected viruses grouped with insect-specific viruses (ISVs) in phylogenetic analyses, suggesting that vertebrates do not participate in the amplification and maintenance cycles of these viruses. However, the novel viruses Cimo peribunyavirus I and II formed a monophyletic clade that shared a most recent common ancestor with viruses of the genera *Orthobunyavirus* and *Pacuvirus*. Orthobunyaviruses are arboviruses that infect a great variety of vertebrates including humans (62). Pacuviruses were isolated from rodents and phlebotomine sandflies in Brazil (75, 76). Thus, the amplification cycle of Cimo peribunyavirus I and II may involve vertebrates and may be more complex than that of the other detected viruses. Further research is necessary to assess the host range of Cimo peribunyavirus I and II. Nevertheless, studying abundance patterns and geographic spread of insect-specific viruses in the light of habitat disturbance and shifts in host community composition can provide valuable insight into infectious disease dynamics. Mosquito-borne viruses are among the major global health

concerns and the recent spread and epidemics caused by Zika and Chikungunya viruses have exemplified that there is an urgent need to understand viral emergence processes (26). Using mosquito-specific viruses as model viruses has many advantages as these viruses are more frequently found than arboviruses and no vertebrate host is required for virus maintenance and amplification which makes links between host community composition and virus abundance patterns easier to identify. Mosquitoes and their viruses are an ideal system to study such effects as this allows studying infectious disease dynamics in an entire family of hosts which can be tested in high numbers.

Our observed virus prevalence patterns are in agreement with several studies stating a heterogeneous effect of biodiversity on pathogen prevalence or disease risk (5, 22). Biodiversity can influence disease risk by different mechanisms like host regulation, changes in encounter rates or transmission rates, leading to an amplifying or diluting net effect (20). Areas with a high host richness likely harbor a high pathogen richness that might act as a source pool of novel diseases upon habitat change and increased contact to humans (2, 3). We observed the highest virus richness in intermediately disturbed and primary rainforest habitats supporting that biodiverse habitats are also rich in pathogens. However, highest viral prevalence rates were observed in villages and at camp sites, further supporting that intact or intermediately disturbed ecosystems also contain well-balanced spectra of viruses.

For those arboviruses, that are transmitted by a generalist mosquito species and use a competent host profiting from disturbance, a dilution effect can occur at local scales (7, 9). Several studies observed a dilution effect for West Nile virus (WNV) with either increasing non-passerine or total bird diversity (11, 77, 78) while others reported no protective effect of avian species richness on WNV prevalence (17, 79). For tick-borne encephalitis virus a dilution effect with increasing density of incompetent deer hosts was observed at local scale (80). Likewise,

the prevalence of the directly transmitted hantavirus sin nombre virus is reduced at sites with higher rodent diversity as the persistence of the main host species (deer mouse) is reduced at diverse sites (12). A similar pattern was observed in our sampling for the five viruses GOLV, HEBV, FERV, Spilikins virus and CAVV belonging to four different families (*Phenuiviridae*, *Peribunyaviridae*, *Phasmaviridae* and *Mesoniviridae*). All these viruses used mosquitoes of the species *Culex nebulosus* as primary host which seemed to profit from disturbance and increased in abundance in disturbed versus pristine habitats (38% vs 3%). The prevalence of the five associated viruses was reduced in the diverse habitat types where the abundance of incompetent mosquito hosts increased and *Culex nebulosus* were rarely found. It is striking that this effect was observed for five taxonomically different viruses infecting the same host species. This effect was not observed for closer related viruses using other mosquito species as host which did not increase in abundance in disturbed habitats. These data suggest that the adaptability of mosquito hosts to changed environments plays a more important role for the increase in prevalence of associated viruses than the phenotypic or genetic characteristics of these viruses.

Contrary effects can be caused by scale-dependent effects or depending on additive or substitutive community assemblies (16). Additionally, the most competent host can either increase or decline with disturbance leading either to a dilution or amplification effect, respectively (7). In our investigated sample set, two viruses (Cimo phenuivirus II, family *Phenuiviridae*, and Cimo flavivirus I, family *Flaviviridae*) with higher prevalence in pristine than in disturbed habitats were detected. As observed for the five viruses mentioned above (GOLV, HEBV, FERV, Spilikins virus and CAVV), virus prevalence rates were closely linked to abundance rates of host mosquito species. Cimo phenuivirus II and Cimo flavivirus I were associated with *Uranotaenia* and *Coquillettidia* mosquitoes that declined in abundance in disturbed habitat

types. An explanation why such an effect is rarely found could be that model pathogens selected in the field of disease ecology are usually zoonotic pathogens that spill over to humans in disturbed habitats and are known to cause outbreaks. Thus, there may exist a bias towards those pathogens that cause problems in disturbed habitat types. Pathogens that are rarely found in disturbed habitats may never be selected for disease ecology studies. These examples underline why an unbiased multi-host and multi-pathogen approach is crucial to uncover host community effects on pathogen prevalence patterns.

In other studies focusing on plant, vector-borne, indirectly and directly transmitted pathogens, the composition of the host community rather than total biodiversity was a predictor for pathogen distribution (5, 21). This is in agreement with the strong observed association between the prevalence of a certain virus and the abundance of its main mosquito host species in our study.

We detected novel strains of two previously characterized viruses, PCLV and Anopheles flavivirus, in *Aedes aegypti* and *Anopheles gambiae* mosquitoes, respectively, corresponding to previous findings of these viruses in these mosquito species in Asia and Africa (60, 66, 81). This further supports the specific associations between certain viruses and mosquitoes on a broader geographical scale.

Mosquitoes harbour a diverse natural virome that can influence subsequent virus infections (82, 83). We observed four frequently co-occurring viruses in *Culex nebulosus* mosquitoes (GOLV, HEBV, FERV and CAVV) that also grew together in cell culture, while in contrast the two *Culex decens*-associated viruses, Cimo rhabdovirus I and JONV, were rarely detected together. This could hint on the one hand at synergistic interactions and on the other hand at superinfection exclusion between different insect-specific viruses in mosquitoes. For several

ISVs in the genera *Flavivirus* and *Alphavirus* an interference with related arboviruses was observed (84-86). A better knowledge of the virome of different mosquito species as well as studies on mutual interference might help to assess vector competence and possibilities of geographic spread.

Virus detection in mosquito homogenates independent of virus isolation extents potential virus findings but has the limitation that exogenous and integrated viruses can be discovered. The detection of likely integrated sequences derived from flavi- and rhabdoviruses in our mosquito sample is in agreement with frequent previous findings of NIRVS from these families in mosquito genomes (87-90). NIRVS are closer related to ISVs than to arboviruses probably limiting their influence on vector competence. The transovarial transmission of ISVs might increase the chance of germline integrations (72, 91). NIRVS can be transcriptionally active and might play a role in immunity against related viruses by producing piRNAs (87, 89, 90, 92).

Collectively, our data show that only some viruses of a huge viral community benefited from ecosystem disturbance. This effect was found for five viruses (GOLV, HEBV, FERV, Spilikins virus and CAVV) and was determined by abundance patterns of their mosquito hosts which profited from habitat disturbance and strongly increased in numbers from pristine to disturbed habitats. In contrast, we also detected two viruses (Cimo phenuivirus II and Cimo flavivirus I) which were associated with mosquito hosts specific to primary habitat types. Detection rates of the associated viruses were also closely linked to host abundance and these viruses were found with higher frequencies in the primary and secondary rainforest. However, our analyses have also shown that virus prevalence is not always a consequence of the abundance of host species in a certain area. The prevalence pattern of JONV did not follow the abundance pattern of *Culex decens* mosquitoes and its abundance seems to be determined by unknown mechanisms, possibly by superinfection exclusion mediated by Cimo rhabdovirus

I. The study of prevalence patterns of a broad genetic diversity of RNA viruses and their associated hosts allowed us to seek for general mechanisms influencing emergence and geographic spread of mosquito-associated viruses. No general dilution or amplification effect was observed. Instead, we identified changes in mosquito community composition causing an increase or decrease of the main host species as the most important factor determining virus prevalence patterns, an effect named abundance effect. Remarkably, host infection rates were not affected by higher host abundance in our study.

Taken together, our data show that host species composition is critical for virus abundance. Environmental changes that lead to an uneven host community composition and to more individuals of a single species is a key driver of virus emergence.

Acknowledgements:

We thank the Ivorian Ministry of Environment and Forest, the Ministry of Research and the directorship of the Taï National Park for the opportunity to conduct this research. We thank all field assistants for help in the field and the Taï chimpanzee project and Fabian Leendertz for logistic support during field work. We are grateful to Verena Heyde, Christian Hieke and Friederike Schröder for excellent laboratory assistance. The work was funded by the Federal Ministry of Education and Research (BMBF) under project number 01KI1716 as part of the Research Network Zoonotic Infectious Diseases and by the Deutsche Forschungsgemeinschaft under project B02 as part of the collaborative research center CRC228.

Author contribution statement:

Conceptualization, S.J. and K.H.; Investigation, K.H., M.M., F.Z. and A.K.; Formal Analysis, S.K.S and K.H.; Resources, S.J.; Supervision, S.J.; Funding Acquisition, S.J.; Writing - Original Draft Preparation, K.H. and S.J.; Writing - Review & Editing, all authors.

Conflict of interest:

The authors declare no conflict of interest.

References:

1. Swei A, Couper LI, Coffey LL, Kapan D, Bennett S. 2019. Patterns, Drivers, and Challenges of Vector-Borne Disease Emergence. *Vector Borne Zoonotic Dis* doi:10.1089/vbz.2018.2432.
2. Dunn RR, Davies TJ, Harris NC, Gavin MC. 2010. Global drivers of human pathogen richness and prevalence. *Proc Biol Sci* 277:2587-95.
3. Keesing F, Belden LK, Daszak P, Dobson A, Harvell CD, Holt RD, Hudson P, Jolles A, Jones KE, Mitchell CE, Myers SS, Bogich T, Ostfeld RS. 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468:647-52.
4. Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646-650.
5. Randolph SE, Dobson AD. 2012. Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology* 139:847-63.
6. Ostfeld RS. 2017. Biodiversity loss and the ecology of infectious disease. *Lancet Planet Health* 1:e2-e3.
7. Halliday FW, Heckman RW, Wilfahrt PA, Mitchell CE. 2017. A multivariate test of disease risk reveals conditions leading to disease amplification. *Proc Biol Sci* 284.
8. Joseph MB, Mihaljevic JR, Orlofske SA, Paull SH. 2013. Does life history mediate changing disease risk when communities disassemble? *Ecol Lett* 16:1405-12.
9. Ostfeld RS, Keesing F. 2012. Effects of Host Diversity on Infectious Disease. *Annual Review of Ecology, Evolution, and Systematics*, Vol 43 43:157-182.
10. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, Jones KE. 2020. Zoonotic host diversity increases in human-dominated ecosystems. *Nature* 584:398-402.
11. Allan BF, Langerhans RB, Ryberg WA, Landesman WJ, Griffin NW, Katz RS, Oberle BJ, Schutzenhofer MR, Smyth KN, de St Maurice A, Clark L, Crooks KR, Hernandez DE, McLean RG, Ostfeld RS, Chase JM. 2009. Ecological correlates of risk and incidence of West Nile virus in the United States. *Oecologia* 158:699-708.
12. Clay CA, Lehmer EM, Jeor SS, Dearing MD. 2009. Sin nombre virus and rodent species diversity: a test of the dilution and amplification hypotheses. *PLoS One* 4:e6467.
13. Ostfeld RS, Keesing F. 2000. Biodiversity and disease risk: The case of lyme disease. *Conservation Biology* 14:722-728.
14. Khalil H, Ecke F, Evander M, Magnusson M, Hornfeldt B. 2016. Declining ecosystem health and the dilution effect. *Sci Rep* 6:31314.
15. Guo F, Bonebrake TC, Gibson L. 2018. Land-Use Change Alters Host and Vector Communities and May Elevate Disease Risk. *Ecohealth* doi:10.1007/s10393-018-1336-3.

16. Johnson PT, Ostfeld RS, Keesing F. 2015. Frontiers in research on biodiversity and disease. *Ecol Lett* 18:1119-33.
17. Levine RS, Hedeem DL, Hedeem MW, Hamer GL, Mead DG, Kitron UD. 2017. Avian species diversity and transmission of West Nile virus in Atlanta, Georgia. *Parasit Vectors* 10:62.
18. Ruyts SC, Ampoorter E, Coipan EC, Baeten L, Heylen D, Sprong H, Matthysen E, Verheyen K. 2016. Diversifying forest communities may change Lyme disease risk: extra dimension to the dilution effect in Europe. *Parasitology* 143:1310-9.
19. Roiz D, Vazquez A, Ruiz S, Tenorio A, Soriguer R, Figuerola J. 2019. Evidence that Passerine Birds Act as Amplifying Hosts for Usutu Virus Circulation. *Ecohealth* doi:10.1007/s10393-019-01441-3.
20. Keesing F, Holt RD, Ostfeld RS. 2006. Effects of species diversity on disease risk. *Ecol Lett* 9:485-98.
21. McLeish M, Sacristan S, Fraile A, Garcia-Arenal F. 2017. Scale dependencies and generalism in host use shape virus prevalence. *Proc Biol Sci* 284.
22. Salkeld DJ, Padgett KA, Jones JH. 2013. A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic. *Ecol Lett* 16:679-86.
23. 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* doi:10.1038/nature20167.
24. Johnson PT, de Roode JC, Fenton A. 2015. Why infectious disease research needs community ecology. *Science* 349:1259504.
25. Gould E, Pettersson J, Higgs S, Charrel R, de Lamballerie X. 2017. Emerging arboviruses: Why today? *One Health* 4:1-13.
26. Weaver SC, Charlier C, Vasilakis N, Lecuit M. 2018. Zika, Chikungunya, and Other Emerging Vector-Borne Viral Diseases. *Annu Rev Med* 69:395-408.
27. Weaver SC, Reisen WK. 2010. Present and future arboviral threats. *Antiviral Res* 85:328-45.
28. Blitvich BJ, Firth AE. 2015. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses* 7:1927-59.
29. Junglen S, Kurth A, Kuehl H, Quan PL, Ellerbrok H, Pauli G, Nitsche A, Nunn C, Rich SM, Lipkin WI, Briesse T, Leendertz FH. 2009. Examining landscape factors influencing relative distribution of mosquito genera and frequency of virus infection. *Ecohealth* 6:239-49.
30. Hermanns K, Zirkel F, Kopp A, Marklewitz M, Rwego IB, Estrada A, Gillespie TR, Drosten C, Junglen S. 2017. Discovery of a novel alphavirus related to Eilat virus. *J Gen Virol* 98:43-49.
31. Hermanns K, Zirkel F, Kurth A, Drosten C, Junglen S. 2014. Cimodo virus belongs to a novel lineage of reoviruses isolated from African mosquitoes. *J Gen Virol* 95:905-9.
32. Junglen S, Kopp A, Kurth A, Pauli G, Ellerbrok H, Leendertz FH. 2009. A new flavivirus and a new vector: characterization of a novel flavivirus isolated from uranotaenia mosquitoes from a tropical rain forest. *J Virol* 83:4462-8.
33. Junglen S, Korries M, Grasse W, Wieseler J, Kopp A, Hermanns K, Leon-Juarez M, Drosten C, Kummerer BM. 2017. Host Range Restriction of Insect-Specific Flaviviruses Occurs at Several Levels of the Viral Life Cycle. *mSphere* 2.
34. Kallies R, Kopp A, Zirkel F, Estrada A, Gillespie TR, Drosten C, Junglen S. 2014. Genetic characterization of goutanap virus, a novel virus related to negevirus, cileviruses and higreviruses. *Viruses* 6:4346-57.
35. Marklewitz M, Handrick S, Grasse W, Kurth A, Lukashev A, Drosten C, Ellerbrok H, Leendertz FH, Pauli G, Junglen S. 2011. Gouleako virus isolated from West African mosquitoes constitutes a proposed novel genus in the family Bunyaviridae. *J Virol* 85:9227-34.
36. 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* 112:7536-41.

37. Marklewitz M, Zirkel F, Rwego IB, Heidemann H, Trippner P, Kurth A, Kallies R, Briesse T, Lipkin WI, Drosten C, Gillespie TR, Junglen S. 2013. Discovery of a unique novel clade of mosquito-associated bunyaviruses. *J Virol* 87:12850-65.
38. Quan PL, Junglen S, Tashmukhamedova A, Conlan S, Hutchison SK, Kurth A, Ellerbrok H, Egholm M, Briesse T, Leendertz FH, Lipkin WI. 2010. Moussa virus: a new member of the Rhabdoviridae family isolated from *Culex decens* mosquitoes in Cote d'Ivoire. *Virus Res* 147:17-24.
39. Zirkel F, Kurth A, Quan PL, Briesse T, Ellerbrok H, Pauli G, Leendertz FH, Lipkin WI, Ziebuhr J, Drosten C, Junglen S. 2011. An insect nidovirus emerging from a primary tropical rainforest. *MBio* 2:e00077-11.
40. Zirkel F, Roth H, Kurth A, Drosten C, Ziebuhr J, Junglen S. 2013. Identification and characterization of genetically divergent members of the newly established family Mesoniviridae. *J Virol* 87:6346-58.
41. Schuster S, Zirkel F, Kurth A, van Cleef KW, Drosten C, van Rij RP, Junglen S. 2014. A unique nodavirus with novel features: mosinovirus expresses two subgenomic RNAs, a capsid gene of unknown origin, and a suppressor of the antiviral RNA interference pathway. *J Virol* 88:13447-59.
42. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3:294-9.
43. Marklewitz M, Dutari LC, Paraskevopoulou S, Page RA, Loaiza JR, Junglen S. 2019. Diverse novel phleboviruses in sandflies from the Panama Canal area, Central Panama. *J Gen Virol* 100:938-949.
44. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson JJ, de Micco P, de Lamballerie X. 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. *J Gen Virol* 85:1971-80.
45. Moureau G, Temmam S, Gonzalez JP, Charrel RN, Grard G, de Lamballerie X. 2007. A real-time RT-PCR method for the universal detection and identification of flaviviruses. *Vector Borne Zoonotic Dis* 7:467-77.
46. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-9.
47. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772-80.
48. Lefort V, Longueville JE, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* 34:2422-2424.
49. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307-21.
50. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* doi:10.1093/molbev/msx335.
51. Hermanns K, Marklewitz M, Zirkel F, Overheul GJ, Page RA, Loaiza JR, Drosten C, van Rij RP, Junglen S. 2019. Agua Salud alphavirus defines a novel lineage of insect-specific alphaviruses discovered in the New World. *J Gen Virol* doi:10.1099/jgv.0.001344.
52. Biggerstaff BJ. 2009. PooledInfRate, Version 4.0: a Microsoft® Office Excel® Add-In to compute prevalence estimates from pooled samples., Centers for Disease Control and Prevention, Fort Collins, CO, U.S.A.
53. R_Core_Team. 2019. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

- 823 54. Chao AG, N. J.; Hsieh, T. C.; Sander, E. L.; Ma, K. H.; Colwell, R. K. and Ellison, A. M. 2014.
824 Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in
825 species diversity studies. *Ecological Monographs* 84:45-67.
- 826 55. Hsieh TCM, K.H. and Chao, A. . 2016. iNEXT: an R package for rarefaction and extrapolation of
827 species diversity (Hill numbers). . *Methods Ecol Evol* 7:1451-1456.
- 828 56. Hsieh TCM, K.H. and Chao, A. 2020. iNEXT: iNterpolation and EXTrapolation for species
829 diversity. R package version 2.0.20.
- 830 57. Legendre PaDC, M. 2013. Beta diversity as the variance of community data: dissimilarity
831 coefficients and partitioning. *Ecology Letters* 16:951-963.
- 832 58. Oksanen JB, F. G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P. R.; O'Hara, R.
833 B.; Simpson, G. L.; Solymos, P.; Stevens, M. H., Szoecs, E. and Wagner, H. 2019. vegan:
834 Community Ecology Package. R package version 2.5-6.
- 835 59. Abudurexiti A, Adkins S, Alioto D, Alkhovsky SV, Avsic-Zupanc T, Ballinger MJ, Bente DA, Beer
836 M, Bergeron E, Blair CD, Briesse T, Buchmeier MJ, Burt FJ, Calisher CH, Chang C, Charrel RN,
837 Choi IR, Clegg JCS, de la Torre JC, de Lamballerie X, Deng F, Di Serio F, Digiario M, Drebot MA,
838 Duan X, Ebihara H, Elbeaino T, Ergunay K, Fulhorst CF, Garrison AR, Gao GF, Gonzalez JJ,
839 Groschup MH, Gunther S, Haenni AL, Hall RA, Hepojoki J, Hewson R, Hu Z, Hughes HR, Jonson
840 MG, Junglen S, Klempa B, Klingstrom J, Kou C, Laenen L, Lambert AJ, Langevin SA, Liu D,
841 Lukashevich IS, et al. 2019. Taxonomy of the order Bunyavirales: update 2019. *Arch Virol*
842 164:1949-1965.
- 843 60. Chandler JA, Thongsripong P, Green A, Kittayapong P, Wilcox BA, Schroth GP, Kapan DD,
844 Bennett SN. 2014. Metagenomic shotgun sequencing of a Bunyavirus in wild-caught *Aedes*
845 *aegypti* from Thailand informs the evolutionary and genomic history of the Phleboviruses.
846 *Virology* 464-465:312-319.
- 847 61. Nault LR, Ammar E. 1989. Leafhopper and Planthopper Transmission of Plant-Viruses. *Annual*
848 *Review of Entomology* 34:503-529.
- 849 62. Holly R. Hughes SA, Sergey Alkhovskiy, Martin Beer, Carol Blair, Charles H. Calisher, Mike
850 Drebot, Amy J. Lambert, William Marciel de Souza, Marco Marklewitz, Márcio R. T. Nunes, Xiǎ
851 ohóng Shí, and the ICTV Report Consortium. 2019. ICTV Virus Taxonomy Profile:
852 *Peribunyaviridae*. *Journal of General Virology* (In Press).
- 853 63. Walker PJ, Blasdel KR, Calisher CH, Dietzgen RG, Kondo H, Kurath G, Longdon B, Stone DM,
854 Tesh RB, Tordo N, Vasilakis N, Whitfield AE, Ictv Report C. 2018. ICTV Virus Taxonomy Profile:
855 *Rhabdoviridae*. *J Gen Virol* 99:447-448.
- 856 64. Valles SM, Chen Y, Firth AE, Guerin DMA, Hashimoto Y, Herrero S, de Miranda JR, Ryabov E,
857 Ictv Report C. 2017. ICTV Virus Taxonomy Profile: Iflaviridae. *J Gen Virol* 98:527-528.
- 858 65. Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, Muerhoff S, Pletnev A, Rico-
859 Hesse R, Smith DB, Stapleton JT, Ictv Report C. 2017. ICTV Virus Taxonomy Profile:
860 *Flaviviridae*. *J Gen Virol* 98:2-3.
- 861 66. Fauver JR, Grubaugh ND, Krajacich BJ, Weger-Lucarelli J, Lakin SM, Fakoli LS, 3rd, Bolay FK,
862 Diclaro JW, 2nd, Dabire KR, Foy BD, Brackney DE, Ebel GD, Stenglein MD. 2016. West African
863 *Anopheles gambiae* mosquitoes harbor a taxonomically diverse virome including new insect-
864 specific flaviviruses, mononegaviruses, and totiviruses. *Virology* 498:288-299.
- 865 67. Attoui H, Mertens, P.P.C., Becnel, J., Belaganahalli, S., Bergoin, M., Brussaard, C.P., Chappell,
866 J.D., Ciarlet, M., del Vas, M., Dermody TS, Dormitzer, P.R., Duncan, R., Fcang, Q., Graham, R.,
867 Guglielmi, K.M., Harding, R.M., Hillman, B., Makkay,, A. M, C., Matthijnsens, J., Milne, R.G.,
868 Mohd Jaafar, F., Mori, H., Noordeloos, A.A., Omura, T., Patton, J.T., Rao,, S. M, M., Stoltz, D.,
869 Suzuki, N., Upadhyaya, N.M., Wei, C. and Zhou, H. 2012. Family Reoviridae, p 541-637. *In*
870 King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), *Virus Taxonomy: classification and*
871 *nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of*
872 *Viruses*. Elsevier Inc., San Diego.
- 873 68. Sadeghi M, Popov V, Guzman H, Phan TG, Vasilakis N, Tesh R, Delwart E. 2017. Genomes of
874 viral isolates derived from different mosquitos species. *Virus Res* 242:49-57.

69. Tomazatos A, Marschang RE, Maranda I, Baum H, Bialonski A, Spinu M, Luhken R, Schmidt-Chanasit J, Cadar D. 2020. Letea Virus: Comparative Genomics and Phylogenetic Analysis of a Novel Reassortant Orbivirus Discovered in Grass Snakes (*Natrix natrix*). *Viruses* 12.
70. Chen R, Mukhopadhyay S, Merits A, Bolling B, Nasar F, Coffey LL, Powers A, Weaver SC, Ictv Report C. 2018. ICTV Virus Taxonomy Profile: Togaviridae. *J Gen Virol* doi:10.1099/jgv.0.001072.
71. Diagne MM, Gaye A, Ndione MHD, Faye M, Fall G, Dieng I, Widen SG, Wood TG, Popov V, Guzman H, Ba Y, Weaver SC, Diallo M, Tesh R, Faye O, Vasilakis N, Sall AA. 2020. Dianke virus: A new mesonivirus species isolated from mosquitoes in Eastern Senegal. *Virus Res* 275:197802.
72. Palatini U, Miesen P, Carballar-Lejarazu R, Ometto L, Rizzo E, Tu Z, van Rij RP, Bonizzoni M. 2017. Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. *BMC Genomics* 18:512.
73. Houe V, Bonizzoni M, Failloux AB. 2019. Endogenous non-retroviral elements in genomes of *Aedes* mosquitoes and vector competence. *Emerg Microbes Infect* 8:542-555.
74. Roiz D, Vazquez A, Seco MP, Tenorio A, Rizzoli A. 2009. Detection of novel insect flavivirus sequences integrated in *Aedes albopictus* (Diptera: Culicidae) in Northern Italy. *Virol J* 6:93.
75. Aitken TH, Woodall JP, De Andrade AH, Bensabath G, Shope RE. 1975. Pacui virus, phlebotomine flies, and small mammals in Brazil: an epidemiological study. *Am J Trop Med Hyg* 24:358-68.
76. Rodrigues DS, Medeiros DB, Rodrigues SG, Martins LC, de Lima CP, de Oliveira LF, de Vasconcelos JM, Da Silva DE, Cardoso JF, da Silva SP, Vianez-Junior JL, Nunes MR, Vasconcelos PF. 2014. Pacui Virus, Rio Preto da Eva Virus, and Tapirape Virus, Three Distinct Viruses within the Family Bunyaviridae. *Genome Announc* 2.
77. Ezenwa VO, Godsey MS, King RJ, Guptill SC. 2006. Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. *Proc Biol Sci* 273:109-17.
78. Swaddle JP, Calos SE. 2008. Increased avian diversity is associated with lower incidence of human West Nile infection: observation of the dilution effect. *PLoS One* 3:e2488.
79. Loss SR, Hamer GL, Walker ED, Ruiz MO, Goldberg TL, Kitron UD, Brawn JD. 2009. Avian host community structure and prevalence of West Nile virus in Chicago, Illinois. *Oecologia* 159:415-24.
80. Cagnacci F, Bolzoni L, Rosa R, Carpi G, Hauffe HC, Valent M, Tagliapietra V, Kazimirova M, Koci J, Stanko M, Lukan M, Henttonen H, Rizzoli A. 2012. Effects of deer density on tick infestation of rodents and the hazard of tick-borne encephalitis. I: empirical assessment. *Int J Parasitol* 42:365-72.
81. Zhang X, Huang S, Jin T, Lin P, Huang Y, Wu C, Peng B, Wei L, Chu H, Wang M, Jia Z, Zhang S, Xie J, Cheng J, Wan C, Zhang R. 2018. Discovery and high prevalence of Phasi Charoen-like virus in field-captured *Aedes aegypti* in South China. *Virology* 523:35-40.
82. Hall RA, Bielefeldt-Ohmann H, McLean BJ, O'Brien CA, Colmant AM, Piyasena TB, Harrison JJ, Newton ND, Barnard RT, Prow NA, Deerrain JM, Mah MG, Hobson-Peters J. 2016. Commensal Viruses of Mosquitoes: Host Restriction, Transmission, and Interaction with Arboviral Pathogens. *Evol Bioinform Online* 12:35-44.
83. Junglen S, Drosten C. 2013. Virus discovery and recent insights into virus diversity in arthropods. *Curr Opin Microbiol* 16:507-13.
84. Nasar F, Erasmus JH, Haddow AD, Tesh RB, Weaver SC. 2015. Eilat virus induces both homologous and heterologous interference. *Virology* 484:51-8.
85. Bolling BG, Olea-Popelka FJ, Eisen L, Moore CG, Blair CD. 2012. Transmission dynamics of an insect-specific flavivirus in a naturally infected *Culex pipiens* laboratory colony and effects of co-infection on vector competence for West Nile virus. *Virology* 427:90-7.
86. Hobson-Peters J, Yam AW, Lu JW, Setoh YX, May FJ, Kurucz N, Walsh S, Prow NA, Davis SS, Weir R, Melville L, Hunt N, Webb RI, Blitvich BJ, Whelan P, Hall RA. 2013. A new insect-specific

- flavivirus from northern Australia suppresses replication of West Nile virus and Murray Valley encephalitis virus in co-infected mosquito cells. PLoS One 8:e56534.
87. Fort P, Albertini A, Van-Hua A, Berthomieu A, Roche S, Delsuc F, Pasteur N, Cappy P, Gaudin Y, Weill M. 2012. Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. Mol Biol Evol 29:381-90.
88. Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. PLoS Genet 6:e1001191.
89. Lequime S, Lambrechts L. 2017. Discovery of flavivirus-derived endogenous viral elements in Anopheles mosquito genomes supports the existence of Anopheles-associated insect-specific flaviviruses. Virus Evol 3:vev035.
90. Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, Oh S, Heiner C, Paxinos E, Andino R. 2017. The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the Aedes aegypti Genome. Curr Biol 27:3511-3519 e7.
91. Olson KE, Bonizzoni M. 2017. Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more? Curr Opin Insect Sci 22:45-53.
92. Ter Horst AM, Nigg JC, Dekker FM, Falk BW. 2019. Endogenous Viral Elements Are Widespread in Arthropod Genomes and Commonly Give Rise to PIWI-Interacting RNAs. J Virol 93.

Table 1: Distribution and host-association of detected viruses. Number of positive pools per habitat and mosquito host species of all detected viruses and virus-like sequences. The main mosquito host species are indicated in bold letters.

Figure legends:

Figure 1: Mosquito species richness per habitat. Rarefaction curves (with 95% confidence intervals) show the observed (solid lines) and interpolated (dotted line) mosquito diversity in the five habitats.

Figure 2: Phylogenetic analyses of detected bunyaviruses. Phylogenetic trees were inferred with PhyML (LG substitution model) based on MAFFT-E protein alignments covering the conserved RdRp motifs of the families *Phenuiviridae* (A), *Peribunyaviridae* (B) and *Phasmaviridae* (C). Novel viruses from this study are indicated in red and previously published viruses detected in our data set are indicated in blue. Sample origin from the different habitat types is indicated by colored circles while no detection is indicated by grey circles. Live virus isolates are marked with a blue virion. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

Figure 3: Phylogenetic analyses of detected rhabdoviruses and iflaviruses. Phylogenetic trees were inferred with PhyML (LG substitution model) based on MAFFT-E protein alignments covering the conserved RdRp motifs of the families *Rhabdoviridae* (A) and *Iflaviridae* (B). Novel viruses from this study are indicated in red and detected virus-like sequences are indicated in grey. Sample origin from the different habitat types is indicated by colored circles while no

detection is indicated by grey circles. Live virus isolates are marked with a blue virion. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

Figure 4: Phylogenetic analyses of detected flaviviruses and orbiviruses. Phylogenetic trees were inferred with PhyML (GTR substitution model) based on MAFFT-E nucleotide alignments covering the conserved RdRp motifs of the genera *Flavivirus* (A) and *Orbivirus* (B). Novel viruses from this study are indicated in red, previously published viruses detected in our data set are indicated in blue and detected virus-like sequences are indicated in grey. Sample origin from the different habitat types is indicated by coloured circles while no detection is indicated by grey circles. Live virus isolates are marked with a blue virion. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

Figure 5: Temperature-dependent replication of novel virus isolates. C6/36 cells were infected with an MOI of 0.1 with Sefomo virus (A), Mikado virus (B), Sassandra virus (C) and Tafomo virus (D). Replication was measured for 3 dpi at 28, 30, 32 and 34°C.

Figure 6: Spearman's rank correlation rho for the most abundant viruses. Only significant correlations are shown. AnthroDist refers to the gradient of anthropogenic disturbance from low to high: PF, primary forest; SF, secondary forest; A, agriculture; C, camp, and V, village.

Figure 7: Richness and cumulative MIR across all tested virus taxa. The number of distinct viruses (A) and the cumulated MIR per 1000 mosquitoes (B) were calculated for all habitat types and for the complete data set. Different virus taxa are shown in different colours.

Figure 8: Prevalence patterns of selected viruses along the disturbance gradient. For all viruses, that were detected in >10 pools, GOLV (A), HEBV (B), FERV (C), Cimo rhabdovirus I (D), CAVV (E), Cimo phenuivirus II (F), Jonchet virus (G), Spilikins virus (H) and Cimo flavivirus I (I), the MIR and MLE per 1000 mosquitoes of the whole data set was calculated for all habitat types. The abundance of the main mosquito host species was plotted. The five viruses GOLV (A), HEBV (B), FERV (C), Cimo rhabdovirus I (D), and CAVV (E) occurred frequently enough in their main mosquito host species (>10 positive pools) to analyze their prevalence in these species. For these viruses, the MIR and MLE per 1000 mosquitoes of the respective species was calculated for all habitat types (right graphs). Significant differences in the infection probability with the most abundant viruses in the different habitats are shown in SI Table 2.

Figure 9: Schematic presentation of the abundance effect. The different habitat types along the anthropogenic disturbance gradient are depicted by photos and drawings. Infection rates are shown for GOLV in all sampled mosquitoes (representing MIR and MLE values as shown in Figure 8) and only in the main mosquito host species, *Culex nebulosus*. The abundance of the main mosquito host species, *Culex nebulosus* is indicated by a grey line.

Supplementary Information:

SI Table 1: Model estimates (base: habitat Primary forest and mosquito group *Anopheles* sp.) for the counts of mosquito individuals per mosquito group and habitat. Significant predictors show the difference to the baseline combination. CulAnn: *Culex annuloris*, CulDec: *Culex decens*, CulNeb: *Culex nebulosus*, Cul_spec: other *Culex* species; Ura_spec: *Uranotaenia* species;

Coq_spec: *Coquilleltidia* species; Ano_spec: *Anopheles* species; others: all other grouped species. PF: primary forest, SF: secondary forest, A: agriculture, C: camp, V: village.

SI Table 2: Significant differences in the infection probability with the most abundant viruses in the different habitats (only Tukey contrasts are shown; bold: $p < 0.05$; italic: $p < 0.1$). No significant differences were detected for FERV ($n = 20$), JONV ($n = 17$), Spilikins virus ($n = 12$) and Cimo flavivirus I ($n = 13$; no findings in camp and village). Combinations with zero virus detections shown in light grey. All virus detections were positively associated with the number of mosquitoes per pool (model estimates not shown).

SI Figure 1: Biodiversity analyses. (A) Hierarchical cluster analysis based on associations between mosquito species. Main mosquito groups were identified as CulAnn, CulDec, CulNeb, Cul_spec, Ura_spec, Ano_spec and Coq_spec. (B) Alluvial plot showing the distribution of the main mosquito species groups to the respective habitats. CulAnn: *Culex annulioris*, CulDec: *Culex decens*, CulNeb: *Culex nebulosus*, Cul_spec: other *Culex* species; Ano_spec: *Anopheles* species; Ura_spec: *Uranotaenia* species; Coq_spec: *Coquilleltidia* species, others: all other grouped species. PF: primary forest, SF: secondary forest, A: agriculture, C: camp, V: village.

SI Figure 2: Phylogenetic analyses of detected phasmaviruses. Phylogenetic trees were inferred with PhyML (WAG (A) and LG (B) substitution model) based on MAFFT-E protein alignments covering the glycoproteins (A) and nucleocapsid (B) proteins of the family *Phasmaviridae*. Novel viruses from this study are indicated in red and previously published viruses detected in our data set are indicated in blue.

SI Figure 3: Phylogenetic analyses of the detected WKOV and members of the genus *Orbivirus*. Phylogenetic trees were inferred with PhyML (LG substitution model) based on MAFFT-E protein alignments of VP3 (A), VP4 (B), VP7 (C) and VP5 (D). WKOV is indicated in red.

SI Figure 4: Phylogenetic analyses of detected reoviruses, alphaviruses and mesoniviruses. Phylogenetic trees were inferred with PhyML (LG substitution model) based on a MAFFT-E protein alignment of the polymerase of members of the family *Reoviridae* (A) or with PhyML (GTR substitution model) based on MAFFT-E nucleotide alignments covering the E2-6K-E1 region of members of the genus *Alphavirus* (B) of the polymerase of members of the family *Mesoniviridae* (C). Novel viruses from this study are indicated in red, previously published viruses detected in our data set are indicated in blue and detected virus-like sequences are indicated in grey. Sample origin from the different habitat types is indicated by colored circles while no detection is indicated by grey circles. Live virus isolates are marked with a blue virion. Abbreviations are PF, primary forest; SF, secondary forest; A, agriculture; C, camp and V, village.

SI Figure 5: Potential NIRVS. (A) PCR amplicons of the generic rhabdovirus PCR assay. RNA/DNA extracts without reverse transcription were used for the PCR. Nested PCR amplicons of Cimo rhabdovirus I and two rhabdovirus-like NIRVS were visualized by ethidium bromide-stained agarose gel electrophoresis. Amplicons with the expected size of 260 bp are framed by a blue box. (B) Schematic representation of selected flavivirus-like NIRVS. Stop codons are indicated by an asterisk, deletions are shown as light grey boxes, frame shifts are indicated by overlapping blue boxes and insertions with similarity to insect genes are shown as waved lines.

SI Figure 6: Cumulative MIR per virus taxon. The cumulated MIR per 1000 mosquitoes was calculated for the analyzed taxa *Phenuiviridae* (A), *Phasmaviridae* (B), *Reoviridae* (C), *Alpha-virus* (D), *Flavivirus* (E), *Peribunyaviridae* (F), *Rhabdoviridae* (G), *Mesoniviridae* (H), and *Iflaviridae* (I), as well as for all detected viruses (J) in the different habitat types and for the complete data set. The different viruses or taxa are shown in different colours.

A

0.2

B

0.5

A **B** **C** **D**

A

B

C

A

Cimo rhabdovirus I (A02)
Rhabdovirus-like NIRVS I (A19)
Rhabdovirus-like NIRVS II (B82)

B

Flavivirus-like NIRVS II (E01)

Flavivirus-like NIRVS III (C78)

Flavivirus-like NIRVS IV (B18)

Flavivirus-like NIRVS X (E11)

A **B** **C** **D** **E** **F** **G** **H** **I** **J**