G3 Genes | Genomes | Genetics

INVESTIGATION

A Survey of Virus Recombination Uncovers Canonical Features of Artificial Chimeras Generated During Deep Sequencing Library Preparation

Jean Peccoud,[1] Sébastian Lequime,[2,3,§] Isabelle Moltini-Conclois,[2,3] Isabelle Giraud,[1] Louis Lambrechts,[2,3] and Clément Gilbert[4]

[1]Laboratoire Ecologie et Biologie des Interactions Unité Mixte de Recherche (UMR) Centre National de la Recherche Scientifique (CNRS) 7267, Université de Poitiers, 86000 France, [2]Insect-Virus Interactions Group, Department of Genomes and Genetics, Institut Pasteur, Paris, France, [3]CNRS, UMR 2000, Paris, France, [4]Laboratoire Evolution, Génomes, Comportement, Écologie, UMR 9191 CNRS, UMR 247 IRD, Université Paris-Sud, 91198 Gif-sur-Yvette, France

ORCID IDs: 0000-0002-3356-7869 (J.P.); 0000-0002-3140-0651 (S.L.); 0000-0001-5958-2138 (L.L.); 0000-0002-2131-7467 (C.G.)

ABSTRACT Chimeric reads can be generated by *in vitro* recombination during the preparation of high-throughput sequencing libraries. Our attempt to detect biological recombination between the genomes of dengue virus (DENV; +ssRNA genome) and its mosquito host using the Illumina Nextera sequencing library preparation kit revealed that most, if not all, detected host–virus chimeras were artificial. Indeed, these chimeras were not more frequent than with control RNA from another species (a pillbug), which was never in contact with DENV RNA prior to the library preparation. The proportion of chimera types merely reflected those of the three species among sequencing reads. Chimeras were frequently characterized by the presence of 1-20 bp microhomology between recombining fragments. Within-species chimeras mostly involved fragments in opposite orientations and located less than 100 bp from each other in the parental genome. We found similar features in published datasets using two other viruses: Ebola virus (EBOV; -ssRNA genome) and a herpesvirus (dsDNA genome), both produced with the Illumina Nextera protocol. These canonical features suggest that artificial chimeras are generated by intra-molecular template switching of the DNA polymerase during the PCR step of the Nextera protocol. Finally, a published Illumina dataset using the Flock House virus (FHV; +ssRNA genome) generated with a protocol preventing artificial recombination revealed the presence of 1-10 bp microhomology motifs in FHV–FHV chimeras, but very few recombining fragments were in opposite orientations. Our analysis uncovered sequence features characterizing recombination breakpoints in short-read sequencing datasets, which can be helpful to evaluate the presence and extent of artificial recombination.

KEYWORDS
artificial chimeras
high-throughput sequencing
Illumina
virus
recombination

Recombination is an important process shaping virus biology and evolution, as it may trigger changes in viral host ranges, virulence and tissue tropism, as well as generate entirely new viruses (Krupovic *et al.*, 2015; Martin *et al.*, 2011; Pérez-Losada *et al.*, 2015; Simon-Loriere and Holmes 2011). In addition to recombining with other viral genomes, viruses may also recombine with the genome of their host, either as an obligate step of their replication cycle, or accidentally (Weiss 2017). Such recombination events may lead to the integration of viral genomes into host genomes, forming endogenous viral elements (EVEs) with varying consequences on host–virus interactions and host biology in general (Feschotte and Gilbert 2012; Katzourakis and Gifford 2010). Reciprocally, virus–host recombination may also result in the integration of a piece of the host genome into a viral genome, supporting the idea that viruses may be efficient vectors of horizontal transfer of genetic material between species (Gilbert and Cordaux 2017).

The present study was initially designed to survey virus–host recombination events between dengue virus (DENV) and *Aedes*

doi: https://doi.org/10.1534/g3.117.300468
Supplemental Material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.10.1534/g3.117.300468/-/DC1.
[§]Current address: Department of Microbiology and Immunology, Rega Institute, KU Leuven-University of Leuven, Leuven, Belgium.
Corresponding author: Clément Gilbert; 1 avenue de la terrasse, Laboratoire Evolution, Génomes, Comportement, Écologie, UMR 9191 CNRS, UMR 247 IRD, Université Paris-Sud, 91198 Gif-sur-Yvette, France; +33 1 69 82 37 37; clement.gilbert@egce.cnrs-gif.fr

albopictus mosquito cells by high-throughput (Illumina) sequencing. Dengue viruses are positive, single-stranded RNA (+ssRNA) viruses belonging to the *Flaviviridae* family (Guzman and Harris 2015). These arthropod-borne viruses are transmitted among humans by mosquitoes of the *Aedes* genus. DENV replication takes place in the cytoplasm of both human and mosquito cells and involves the production of a double-stranded RNA intermediate generated by the virus-encoded RNA-dependent RNA-polymerase (RdRp) (Bartenschlager and Miller 2008). The frequent finding of defective DENV genomes resulting from virus–virus non-homologous recombination in DENV-infected patients (Li *et al.*, 2011) led us to hypothesize that DENV may also recombine at an appreciable frequency with RNAs of their host (human or mosquito).

Rather than biological chimeras, our analysis detected a large number of artificial chimeras. Two of the enzymes involved in protocols of sequencing library preparation are known to generate such chimeras. Reverse transcriptases used to generate cDNA libraries are known to be prone to both inter- and intra-molecular template switching (Cocquet *et al.*, 2006; Zeng and Wang 2002). DNA polymerases used in some library preparation protocols can also generate *in vitro* recombination, via pausing on an incompletely elongated strand or premature termination, and reannealing of the strand to another DNA fragment in the following PCR cycle (Meyerhans *et al.*, 1990; Pääbo *et al.*, 1990). Such biases, as well as their associated problems in many areas of research, have long been recognized and various protocol optimizations and controls have been proposed to take them into account (Di Giallonardo *et al.*, 2013; Edgar *et al.*, 2011; Görzer *et al.*, 2010; Haas *et al.*, 2011; Routh *et al.*, 2015; Shao *et al.*, 2013; Zanini *et al.*, 2017). In this study, we were not able to observe any convincing evidence of biological recombination between DENV and mosquito RNA, suggesting that it is unlikely to be frequent. However, using appropriate controls, we investigated the nature of artificial chimeras detected in our sequencing libraries. To generalize our findings, we also analyzed multiple published sequence datasets generated with similar sequencing protocols. Together, our results uncover sequence features typical of artificial chimeras, which may help detect and characterize such artifacts in future studies.

MATERIAL AND METHODS

Virus and cell lines

C6/36 (*Aedes albopictus*) mosquito cells were maintained at 28° in Leibovitz's L-15 medium (Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies), 1% non-essential amino acids (Life Technologies) and 0.1% Penicillin-Streptomycin (Life Technologies) as previously described (Fontaine *et al.*, 2016). A sub-confluent culture of C6/36 cells grown in a 75-cm^2 flask was inoculated with DENV type 1 isolate KDH0026A (Fansiri *et al.*, 2013) at a multiplicity of infection of 0.01 as previously described (Fontaine *et al.*, 2016). The cell culture supernatant was harvested after 5 days of incubation at 28°.

Library preparation

Total RNA was extracted from the supernatant of the DENV-infected C6/36 cell culture using TRIzol (Life Technologies). Total RNA was treated with Turbo DNAse (Life Technologies) and purified with magnetic beads (Beckman Coulter). Several studies have shown that artificial recombination events can occur at elevated rates during the RT-PCR and/or PCR steps involved in the preparation of sequencing libraries. To subsequently estimate the amount of such technical chimeras, mosquito + DENV RNA was mixed with purified RNA from the common pillbug *Armadillidium vulgare*. Given that the analyzed pillbug and DENV have never been in contact before the mix, any junction identified between their RNAs can only be artificial. Pillbug RNA was used because we had easy access to this material and because pillbugs and mosquitoes diverged ~500 million years ago, ensuring good confidence in the assignment of chimeric sequencing reads to one or the other species. The mix was performed to target a 1:1 ratio in the concentrations of the two RNA samples, following RNA quantification by fluorometry (Qubit RNA HS, Life Technologies).

The RNA mix was reverse transcribed in duplicates using Transcriptor High Fidelity cDNA Synthesis Kit (Roche Applied Science) and random hexamers. The second strand was synthetized in a single reaction with *E. coli* DNA ligase (New England Biolabs [NEB]), *E. coli* DNA polymerase I (NEB), *E. coli* RNAse H (NEB) in second-strand synthesis buffer (NEB). The newly synthesized dsDNA was purified with magnetic beads (Agencourt AMPure XP, Beckman Coulter) and its concentration was measured by fluorometric quantification (Quant-iT PicoGreen dsDNA, Invitrogen).

A sequencing library was prepared in duplicate using Nextera XT DNA Library Preparation Kit (Illumina) following the manufacturer's instructions. Duplicate libraries were multiplexed with other libraries and sequenced in single end on an Illumina NextSequation 500 platform using a mid-output 150 cycles v2 kit (Illumina). Sequencing reads were demultiplexed using bcl2fastq v2.15.0 (Illumina).

Detection of chimeric reads

Detection of chimeric reads resulting from either host–virus or virus–virus recombination used an approach previously developed to search for moth sequences integrated into genomes of the AcMNPV baculovirus (Gilbert *et al.*, 2016). Briefly, this approach involves blastn (Camacho *et al.*, 2009) homology searches of the sequencing reads against reference genomes and/or transcriptomes. Searches were performed separately with the DENV genome (Genbank accession number: HG316481), with the *Ae. albopictus* genome (MNAF02000000) and with the *Ar. vulgare* genome (LYUU01000000), allowing for the recovery of two possible hits per read in each blast (option –max_target_seqs 2).

The following steps were performed with custom R (R Core Team 2017) scripts. All reads containing a region of at least 20 bp that aligns multiple times to one or more reference genomes were identified. Among those alignments, only the one with the best blast score was kept. A read was then identified as chimeric when two portions align to two genomes or two non-adjacent regions of the same genome (in the following only the former case is considered for simplicity, but the principles apply to both). Each region of such read must align over at least 28 bp, and over at least 16 bp to one genome *only*. The read must be aligned to both genomes over at least 90% of its full length. Aligned read regions are allowed to overlap by up to 20 bp or to be separated by at most 5 bp. The overlap corresponds to a microhomology between both genomes at the recombination point, and the separation should correspond to non-templated addition of nucleotides at this point. These criteria were based on the distribution of real data in Gilbert *et al.* (2016). Reanalysis of the datasets published by Routh *et al.* (2015) (see below) yielded chimeric read counts similar to those obtained by these authors despite the use of different methods. Importantly, the approach designed by Routh and Johnson (2014) and Routh *et al.* (2015) does not allow the identification of overlapping alignments in the chimeric reads. It first uses a read mapper [Bowtie2 (Langmead and Salzberg 2012)] to find the best alignment between a chimeric read and a reference genome, cuts the remaining unaligned portion of the chimeric read and searches for the best alignment between this second read portion and the same or another reference genome. Hence, the

second alignment cannot include a portion of the read involved in the alignment of the first read portion, thereby precluding recovery of possible overlaps between alignments of the two read portions.

To check whether microhomology lengths at recombination points were consistent with those expected by chance, considering the sequences of the analyzed genomes, the distribution of homology length was compared to that of random chimeric reads generated *in silico*. Each *in silico* read was made of two regions extracted from random locations of a pair of genomes (DENV, EBOV, herpesvirus, FHV or pillbug RNA genes). The lengths of the two regions were chosen at random, with the conditions that both were at least 28 bp and their sum was the size of a read (150 bp). These reads were then blasted against the sequences from which they were generated and the blast outputs were submitted to the same analysis as that performed on real data.

Chimeric reads potentially resulting from PCR duplicates generated during the preparation of the Illumina sequencing library were identified as reads having the same alignment coordinates in the two genome sequences they involve. Only one read among duplicates was retained. This selection leads to conservative estimates of the number of chimeras, at it is not excluded that apparent PCR duplicates result from sequencing the exact same portion (same start and end position) of distinct RNA molecules generated by non-PCR (viral) replication of a host–virus or virus–virus recombination event. The probability of sequencing the same original biological chimera at the same position increases with sequencing depth.

Annotation of transposable elements and ribosomal RNA genes

Many transposable elements (TEs) from a host (moth) genome have been shown to transpose into the genome of a DNA virus (the AcMNPV) (Fraser *et al.*, 1983; Gilbert *et al.*, 2016; Miller and Miller 1982). TEs indeed represented virtually all the moth sequences that were linked to AcMNPV sequences and they were inserted at their preferred target duplication sites in the virus genome (Gilbert *et al.*, 2016). Here, it was not expected to find signs of *bona fide* transposition of mosquito TEs into DENV genomes because TEs transpose into DNA, not into RNA, and the DENV replication cycle does not normally involve reverse transcription of the DENV genome into DNA. Yet, it has been shown that upon infection of flies and mosquitoes, a number of RNA viruses including DENV produce viral DNA (vDNA) that is transcribed and involved in small-RNA-mediated antiviral immunity (Goic *et al.*, 2013, 2016). Importantly, these vDNAs are generated by template switching of reverse transcriptases encoded by host retroelements. As a result, they are frequently flanked by retrotransposon sequences. It is possible that transcription of vDNAs could initiate or terminate in their upstream or downstream TE flanking sequence, which could produce TE-virus chimeric transcripts. In this context, the production of vDNAs could increase the likelihood of TEs, compared to non-TE sequences, to be joined to viral genome fragments and to be encapsidated into virions. To assess whether mosquito–DENV chimeric reads involve mosquito TEs, the alignment positions of chimeric reads were located with respect to TE copies that were annotated by Peccoud *et al.* (2017).

Ribosomal RNA genes were also annotated in the mosquito and pillbug genomes. For this, we used a full-length copy of *Ae. albopictus* ribosomal RNA genes (accession number: L22060) and a set of Isopoda partial RNA genes (Supplementary Table 1) as queries to perform blastn searches with default parameters on the genome of the two species. No annotated full-length copy of all ribosomal RNA genes is available for *Ar. vulgare*.

Analysis of other types of chimeras

To better understand the process of chimera formation, other types of chimeric reads were analyzed. Because DENV–DENV chimeric reads involve the same, very short, genome, the following properties could be assessed: (i) the distance between recombining sequences in the source genome, relative to the distance distribution between two randomly drawn positions in this genome, (ii) the number of breakpoints detected in 100-bp windows sliding along the DENV genome, and (iii) the orientation of the two recombining sequences (same or different orientations) in each chimera. Pillbug–pillbug chimeras were investigated in the same fashion. This analysis focused on chimeras involving two sequences of 18S ribosomal pillbug RNA, which composed a large number of pillbug–pillbug chimeras. Ribosomal 18S RNA chimeras were identified by blasting all reads on an 18S rRNA sequence of *Ar. vulgare* (Genbank accession number: AJ287061.1).

The presence and properties of chimeras found in other datasets was examined in a published high-throughput sequencing dataset of EBOV produced with the same library preparation protocol as in the present study (Gire *et al.*, 2014). Unlike DENV, EBOV (*Mononegavirales*, *Filoviridae*) is a -ssRNA virus. To detect EBOV–EBOV chimeras, all reads were aligned to the Makona-G3686 EBOV genome (accession number: KM034562) with blastn, followed by the same analysis of blast results described above. Whether chimeras could be formed during the synthesis of cDNA or during the PCR step of the Nextera library-preparation protocol was assessed by analyzing in the same fashion a published Illumina dataset produced to sequence the genome of a DNA virus, the Macropodid herpesvirus 1 (MaHV-1) (Vaz *et al.*, 2016).

Finally, a high-throughput sequencing dataset of FHV published by Routh *et al.* (2015), which is devoid of artificial chimeras, was reanalyzed. This dataset allowed the characterization of properties of biological recombination events in an RNA virus, in particular the microhomology length between recombining sequences, using the same protocol applied to the other datasets. In their study, Routh *et al.* (2015) used click chemistry rather than enzymatic reactions for the ligation of Illumina adaptors. The method generates unbiased libraries, with an extremely low artificial recombination rate. FHV is a +ssRNA virus belonging to the *Alphanodavirus* genus in the *Nodaviridae* family. It encapsidates a bipartite genome, composed of RNA1 (3,107 bp), which encodes the viral RdRp, and RNA2 (1,400 bp), which encodes the viral capsid protein.

Data availability

Raw sequence reads data from DENV, *Ae. albopictus* and *Ar. vulgare* are available in fastq format at the NCBI short read archive under accession number SRP129541. The R codes used for the data analysis are provided in File S1.

RESULTS

Biological recombination Between DENV and host RNA is likely infrequent

Alignments of the reads to the DENV genome and to the *Ae. albopictus* mosquito and *Ar. vulgare* pillbug genomes yielded similar numbers of aligned reads for the two technical replicates that we performed (Table 1). The ratio of DENV + mosquito reads over pillbug reads were 2.6 (replicate 1) and 3.4 (replicate 2). The average sequencing depth of the DENV genome was 26,423X (replicate 1) and 28,594X (replicate 2). Hereafter, we refer to total numbers of reads obtained in the two replicates.

We identified 3,717 chimeric reads containing a junction between pillbug and DENV RNA. These reads represent 3,639 breakpoints that all differ from each other by their locations (Table 1) and which must have been created *in vitro* since pillbug and DENV RNAs have not been in contact *in vivo*. The observed proportion of pillbug–DENV chimeric

Table 1 Number of chimeric reads detected in all datasets analyzed in this study

Type of chimeras	Total number of chimeras	Proportion of chimeras*	Number of unique chimeras	Total replicates 1 + 2
DENV 1–Moquito 1	350	0.013–0.13%	343	555
DENV 2–Mosquito 2	216	0.007–0.098%	212	
DENV 1–Pillbug 1	2353	0.09–0.22%	2304	3639
DENV 2–Pillbug 2	1364	0.05–0.16%	1335	
DENV 1–DENV 1	212644	8%	44805	81516
DENV 2–DENV 2	189633	6.70%	36711	
EBOV–EBOV	144556	7.21%	16504	—
Pillbug18S–Pillbug18S 1	20324	4.90%	4662	817
Pillbug18S–Pillbug18S 2	14799	4.56%	3508	
MaHV-1–MaHV-1	2507	0.12%	1139	—
FHV-RNA1–FHV-RNA1	744201	2.90%	3675	—
FHV-RNA2–FHV-RNA2	45155	0.61%	1313	—

* The percentage of total reads made up by chimeras. For inter-genome chimeras, the proportions are given in respect of the total number of reads mapping on each genome.

reads involving pillbug ribosomal RNA (73%) was not significantly different from the proportion of all pillbug reads assigned to rRNA (Chi^2 *p-value* = 0.4; 1 df). This indicates that the probability for pillbug sequences to recombine *in vitro* with DENV sequences mainly depends on their relative proportions in the RNA extracts. The proportion of pillbug and mosquito reads aligning to rRNA genes (74% and 77%, respectively) is consistent with previous RNAseq datasets produced without rRNA depletion (O'Neil *et al.*, 2013).

We detected 566 mosquito–DENV chimeric reads, representing 555 unique chimeras. None of the recombination breakpoints we sequenced were found to involve a mosquito TE. Instead, most of the chimeras (74%) contain mosquito rRNA, a proportion that is not significantly different from that observed among all mosquito reads (Chi^2 *p-value* = 0.4; 1 df). It is noteworthy that the number of mosquito–DENV chimeras is lower than expected given the number of pillbug–DENV chimeras and the proportion of reads from each species in the whole dataset (Chi^2 *p-value* < 0.001; 1 df). The reason underlying this discrepancy is unclear and we could not find any evidence suggesting it may be biological. On the opposite, the lower proportion and similar distribution of microhomology length of mosquito-DENV chimeras compared to artificial pillbug-DENV chimeras (Figure 1), indicate that most, if not all, mosquito-DENV chimeras were artificially produced.

Interestingly, the majority of the pillbug–DENV and mosquito–DENV junctions (56% and 57%, respectively) are characterized by the presence of 1-20 bp microhomologies (Figure 1). The distributions of microhomology lengths for both types of chimeras are very similar to each other and at the same time clearly higher than those expected by chance. This shows that linked sequences tend to be similar at their recombination point. Still, a substantial proportion of pillbug–DENV (26.7%) and mosquito–DENV (28%) chimeras devoid of microhomologies are characterized by the presence of 1-5 nucleotides that did not derive from any of the two recombining sequences. The presence of these additional nucleotides is likely due to the known ability of DNA polymerases to synthesize DNA in a non-template directed manner (Clark 1988; García *et al.*, 2004).

Canonical features of chimeras found in various Illumina Nextera sequencing datasets

DENV–DENV chimeric reads were far more numerous than arthropod–DENV ones (Table 1), which can be explained by the high proportion (70%) of DENV reads in the datasets. DENV–DENV breakpoints are also characterized by the presence of 1–20 bp microhomologies and the microhomology length distribution clearly differs from that expected by chance (Figure 2A). Most DENV–DENV recombination events involved sequences separated by distances that are much shorter than expected if recombination took place randomly along the DENV genome (Figure 2B). Plotting the density of recombination breakpoints along the DENV genome (Figure 2C) did not reveal any particular hotspots, indicating that recombination takes place along the entire length of the DENV genome at similar frequencies. Interestingly, the vast majority of DENV–DENV chimeras (60,781 out of 81,516, ~74.5%) involve sequences in opposite orientation.

We detected 8,170 unique intra-pillbug ribosomal RNA breakpoints that are characterized by the same features as DENV–DENV breakpoints, *i.e.*, most of them (81%) show 1–20 bp microhomology as well as evidence of non-templated nucleotide addition and breakpoints devoid of microhomology (Supplementary Figure 1), recombination between fragments in opposite orientation (84.4% of the chimeras) that are typically distant by a few dozen nucleotides.

By analyzing the EBOV dataset of (Gire *et al.*, 2014), we found 144,556 chimeric reads representing 16,504 different EBOV–EBOV breakpoints, which were characterized by the same features as DENV–DENV or pillbug–pillbug chimeras (Supplementary Figure 2). The length of microhomology between recombining sequences was longer than expected by chance, most recombining sequences were distant from less than 100bp and were in opposite orientations (65% and 92%, respectively).

Finally, our analysis of the herpesvirus dataset of Vaz *et al.* (2016) detected 2,507 MaHV-1–MaHV-1 chimeric reads representing 1,139 different breakpoints. These were characterized by features similar to those of RNA breakpoints (Supplementary Figure 3): 95% show 1–20 bp microhomology and recombination occurs predominantly (at 93%) between nearby fragments in opposite orientations.

Features of natural recombination Between viral sequences

In the FHV dataset of Routh *et al.* (2015), we found a total of 744,201 FHV-RNA1–FHV-RNA1 and 45,155 FHV-RNA2–FHV-RNA2 chimeric reads representing respectively 3,675 and 1,313 unique recombination breakpoints. We note that in spite of the differences between our approach and that of Routh *et al.* (2015) in terms of alignment method (mapping *vs.* blast) and criteria to detect chimeric reads, our numbers are similar to those obtained in the original study (766,143 FHV-RNA1–FHV-RNA1 and 39,856 FHV-RNA2–FHV-RNA2;

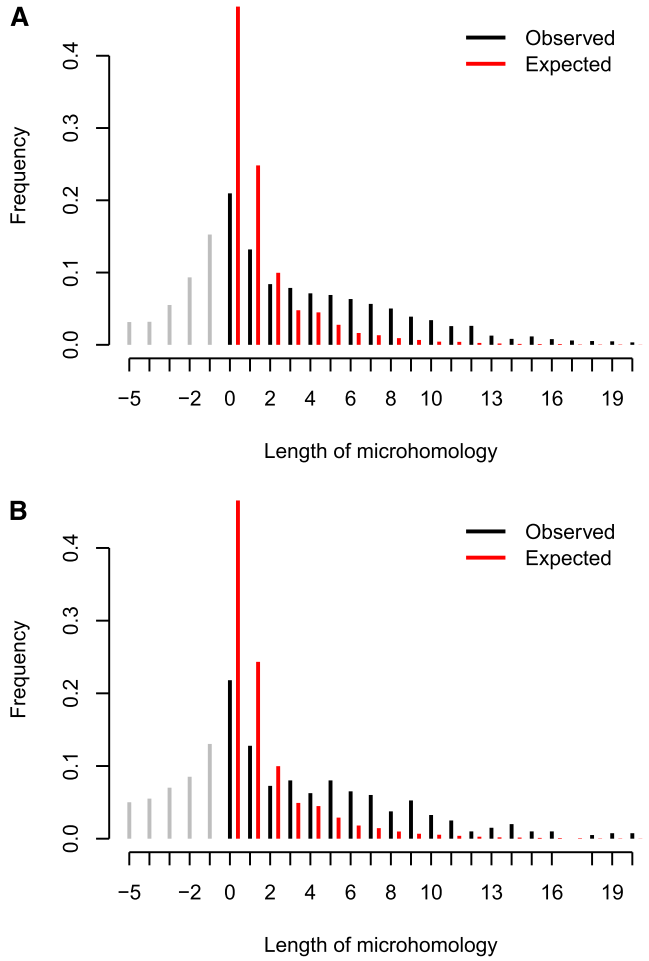

Figure 1 Distribution of lengths of microhomology motifs found in mosquito–DENV (A) and pillbug–DENV (B) breakpoints. The black/gray and red histograms show the observed and expected distributions, respectively. Negative lengths (gray bars) represent insertion of non-templated nucleotides at junction points. Frequencies of observed microhomologies were rescaled so that the heights of black bars sum to 1. This was needed for comparison to the expected distribution, which does not account for negative microhomology.

Routh *et al.* 2015). Among these, very few breakpoints involved genome regions in opposite orientation (73 out of 4,988, ~1.5%), indicating that recombination occurred via a mechanism different from that underlying DENV–DENV, EBOV–EBOV and MaHV-1–MaHV-1 chimeras. Interestingly, we found that the majority (64%) of FHV–FHV breakpoints were characterized by the presence of 1-20 bp microhomologies, with the bulk (89%) of these microhomologies being 1-7 bp (Figure 3A). The distribution of distances between sequences involved in chimeras was wider than those previously described, but still narrower than that expected by chance (Figure 3B).

DISCUSSION

Formation of artificial chimeras During PCR

We evaluated the biological origin of host-virus (mosquito-DENV) recombination by comparing its rate to that of recombination between genomes (pillbug-DENV) that never were in contact in a cellular context. This control indicated that the DENV genome and the genome of its mosquito host did not recombine *in vivo* to a detectable extent, hence, that most, if not all, virus-host chimeras were artificially produced.

We also uncovered within-genome chimeras (*e.g.*, DENV-DENV chimeras) in greater numbers. To assess their biological origin, chimeras involving the genomes of species that were not in contact prior to DNA processing (*e.g.*, pillbug-DENV chimeras) cannot be used as a baseline expectation because they necessarily occurred between different molecules, contrary to within genome chimeras, which most frequently occurred within the same molecule (see below). Thus, they are not informative. However, it is notable that the various within-genome chimeras we detected in datasets obtained with the Illumina Nextera library preparation kit have similar features, whether they involve RNA or DNA templates from viruses or other source material (*e.g.*, pillbug): (i) recombining fragments present comparable microhomology length distributions at their junction points, (ii) junction sites appear evenly distributed along molecules, (iii) distances between recombining regions of are mostly shorter than 100 pb, and (iv) chimeric reads mostly link sequences in opposite orientations. These features were not shared by FHV chimeras sequenced by the ClickSeq technology (Routh *et al.*, 2015), whose biological origin is established. Hence, the sequencing technology, rather than differences between analyzed organisms, seems to explain features of chimeric reads. We therefore conservatively interpret Nextera-sequenced chimeras as produced during the processing of genetic material for sequencing.

Because sequencing of the DNA virus MaHV-1 did not require cDNA synthesis by reverse transcription, recombination events detected in this dataset must have taken place during the PCR step of the Nextera library preparation protocol. Short recombination distances indicate that most breakpoints are generated by intra-molecular, rather than inter-molecular, recombination between close positions. Indeed, there is no reason for nearby sequences to preferentially recombine if these belonged to different molecules, considering the absence of recombination hotspots along genomes. This observation also excludes chimera production during the sequencing reaction. Such error may have arisen if the fluorescence signals generated by sequencing two different DNA fragments were somehow swapped in the course of the run. Such a reading error has no reason to preferentially involve fragments coming from nearby locations in the source genome and in opposite orientation. As for chimeras formed during the bridge PCR generating a DNA cluster on the flow cell, these are expected to generate garbled reads, not chimeric reads.

Various models, most of which are experimentally verified, have been proposed to explain the formation of chimeras during PCR (Kanagawa 2003; Shammas *et al.*, 2001). These include several types of template-switching mechanisms (Guieysse *et al.*, 1995; Odelberg *et al.*, 1995; Patel *et al.*, 1996). The majority of the chimeric reads detected in our study show the very features expected to be produced by template switching during extension of a primer along a double-stranded (or partially double-stranded) template (Guieysse *et al.*, 1995; Patel *et al.*, 1996) (Figure 4). In this configuration, strand displacement is catalyzed by the polymerase. Primer extension may proceed poorly as the displaced strand tends to reanneal to the template, inducing dissociation of the extending strand. The dissociated extending strand may then reanneal to the displaced strand at a position downstream of the primer binding site (Patel *et al.*, 1996). Reannealing with the displaced strand can explain why the two fragments composing a chimera tend to have opposite orientations. Our data also show that chimeras involve an excess of microhomology of a few nucleotides. This suggests that the 3′ end of the extending strand often reanneals over a few complementary bases on the displaced stand. Consistent with this deduction, template switching was shown to be most efficient when annealing occurred over 7-9 bp, in some PCR conditions (Guieysse *et al.*, 1995). As few as three adjacent complementary base pairs appear sufficient to favor template switching

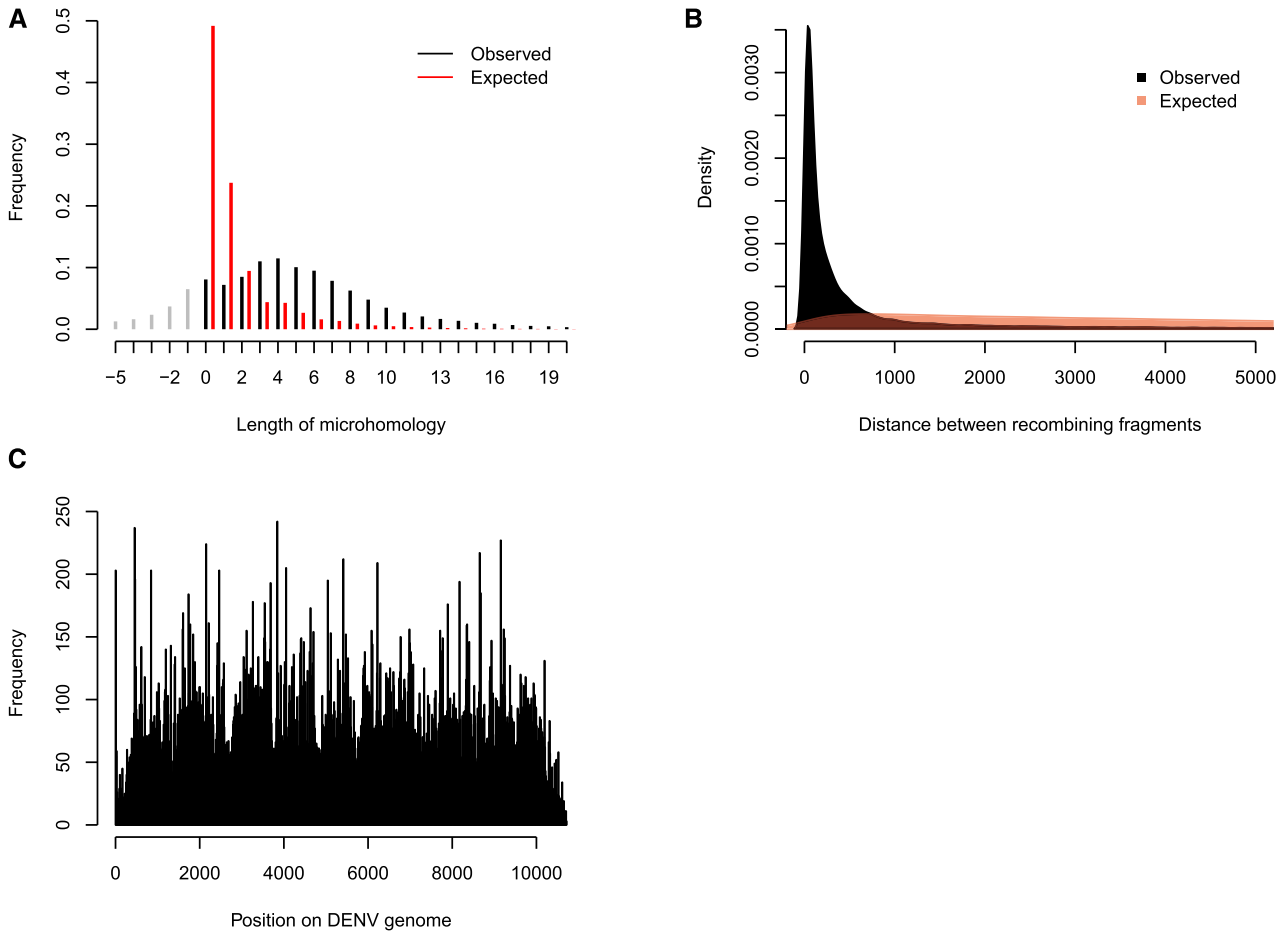

Figure 2 Characteristics of DENV–DENV breakpoints. (A) Distribution of lengths of microhomology motifs, as in Figure 1. Negative lengths (gray bars) represent insertion of non-templated nucleotides at junction points. Frequencies of observed microhomologies were rescaled so that the heights of black bars sum to 1. (B) Density of observed (black/gray) and expected (red) distances separating recombining fragments found in DENV–DENV himeras. (C) Number of DENV–DENV breakpoints per 100-bp, non-overlapping windows along the DENV genome.

during elongation, since microhomologies of 3 pb and longer are more frequent than expected (Figures 1-3), but we cannot evaluate whether shorter lengths (1-2 pb) promote template switching.

We could not find any published experimental work that would help explain why recombining sequences are preferentially located at short distances from each other (<100 bp). We suspect this pattern of preferred distance might be due to steric constraints inherent to the polymerase and/or to other PCR conditions, including the size of the template. Finally, many of the chimeras do not show the features expected to result from the intra-molecular template-switching mechanism characterized by Patel *et al.* (1996), including those showing no microhomology, non-templated addition of nucleotides or involving different source genomes, hence different molecules (like arthropod–DENV chimeras). While many of these chimeras may have been generated through other, unknown, *in vitro* recombination mechanisms, the possibility remains that some of them are biological chimeras. Nevertheless, the robust detection and analysis of biological recombination in RNA viruses clearly requires new library preparation protocols that suppress or drastically reduce the rate of artificial recombination, such as the one developed by Routh *et al.* (2015) and Jaworski and Routh (2018).

Remarkably, the features characterizing artificial chimeras produced with the PCR-based Illumina Nextera library preparation protocol are very similar to those generated by the non-PCR-based, Multiple Displacement Amplification (MDA) protocol routinely used for whole-genome amplification (Lasken and Stockwell 2007). An analysis of 475 artificial chimeras generated by MDA followed by 454 (Life Sciences) sequencing revealed a majority of fragments in opposite orientation (85%) and the presence of 2-21 bp microhomologies in 70% of all MDA chimeras (Lasken and Stockwell 2007). The MDA differs from the PCR in several aspects, such as the use of random hexamer primers, the absence of cycles of varying temperatures and the production of branched, long DNA molecules. Yet, the mechanism thought to be responsible for the formation of MDA chimeras is virtually identical to the one explaining PCR-based chimeras, and may perhaps even be reinforced by the strong displacement activity of the phi29 DNA polymerase used in MDA (Lasken and Stockwell 2007).

Biological recombination of RNA viruses may be facilitated by microhomology

Most FHV sequences generated using the ClickSeq library preparation protocol (Routh *et al.*, 2015) and involved in chimeras presented 1-7 bp microhomologies, which is longer than expected by chance (Figure 3). This is fully consistent with an earlier low-throughput study that found that 24 out of 40 FHV–FHV breakpoints were characterized by the presence of 1-7 bp microhomology, supporting an occasional role for

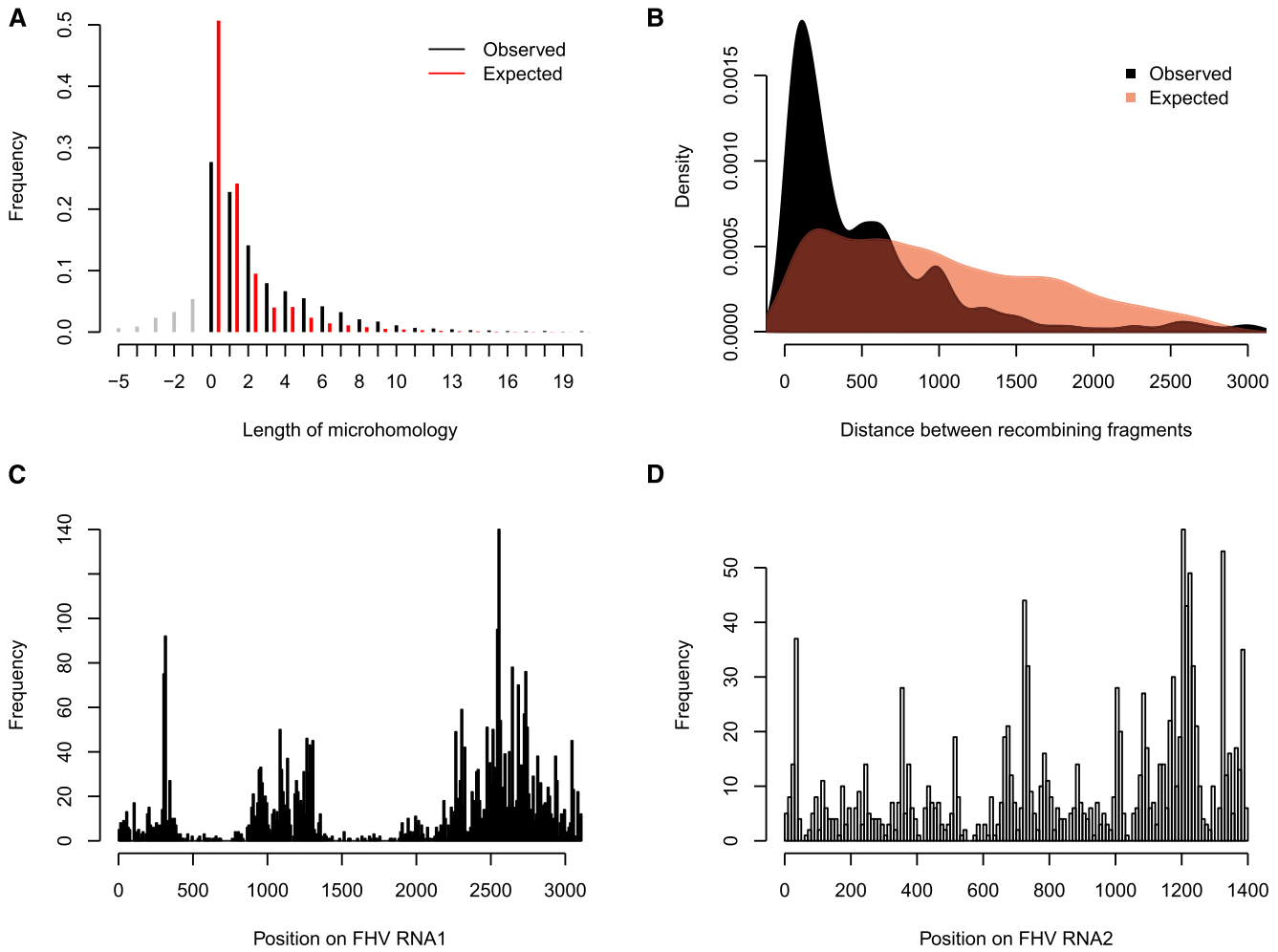

Figure 3 Characteristics of FHV–FHV breakpoints. (A) Distribution of lengths of microhomology motifs as in Figure 1. Negative lengths (gray bars) represent insertion of non-templated nucleotides at junction points. Frequencies of observed microhomologies were rescaled so that the heights of black bars sum to 1. (B) Density of observed (black/gray) and expected (red) distances separating recombining fragments found in FHV–FHV chimeras. The number of FHV–FHV breakpoints per 100-bp, non-overlapping windows along the FHV genome is shown in (C) for the FHV RNA1 and (D) for the FHV RNA2.

base pairing between the nascent and template strand in facilitating recombination (Li and Ball 1993). Much like Li & Ball (1993), we found a substantial number of FHV–FHV breakpoints devoid of microhomology, suggesting that base pairing is not necessary for recombination to occur.

The mechanisms inducing RNA virus recombination can broadly be classified into two categories, replicase-driven template switching and RNA-breakage induced recombination (Nagy and Simon 1997). To our knowledge, FHV recombination has never been studied experimentally. However, the experimental study of defective interfering RNAs from other RNA viruses such as Tombusvirus (White and Morris 1995) and Bromovirus (Pogany *et al.*, 1997) suggests that the presence of microhomology is in agreement with recombination taking place via intramolecular template switching of the viral RdRp (Nagy and Simon 1997). A fairly large number of FHV–FHV chimeras were characterized by the presence of non-templated addition of nucleotides (554 out of 4,988). Such non-templated nucleotides have previously been found at recombination breakpoints of the turnip crinkle virus (Cascone *et al.*, 1990), tobamovirus (Raffo and Dawson 1991) and brome mosaic virus (Nagy and Bujarski 1996). Interestingly, the addition of such nucleotides by the RdRp at the 3′ end of the nascent strand has been proposed as one of the factors triggering partial dissociation of the nascent strand from the first template, pausing of the RdRp and template switching (Nagy and Simon 1997). In terms of distance between recombining regions, we found a pattern partially mirroring that generated by artificial recombination, with an enrichment of template switches over short distances ($<$ 100 bp) (Figure 3B). However, the enrichment toward short distances was clearly less pronounced than for artificial recombination, and we observed two other marked peaks of longer preferred distances ($>$ 500 bp). Another characteristic differentiating artificial chimeras from FHV-FHV chimeras is the presence of hotspots of recombination along the viral genome (Figure 3C,D), which were previously detected by Routh and Jonhson (2014).

Conclusion

This study was initiated to search for evidence of *in vivo* host-RNA virus recombination in the DENV–mosquito cell system. The overrepresentation of technical chimeras resulting from *in vitro* recombination prevented us from identifying any such event, be it between host and DENV or intra-DENV. We found no evidence in support of frequent

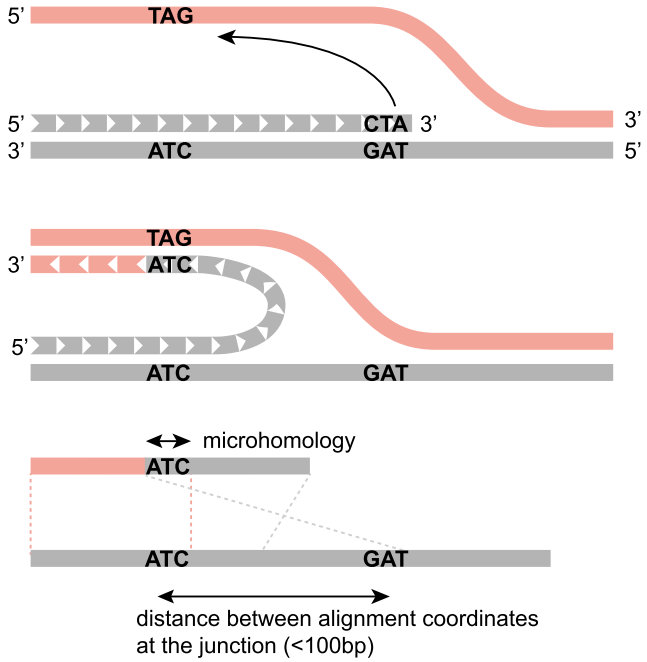

Figure 4 Model for the formation of artificial chimeras during the PCR step of the Illumina Nextera library preparation kit. Artificial chimeras most likely take place by intra-molecular template switching (TS) during extension of a primer along a double-stranded (or partially double-stranded) template. The dissociated extending strand reanneals to the displaced strand at a position upstream in respect to the direction of extension. Such reannealing is facilitated by base pairing over 1-20 bp (here, 3 bp), *i.e.*, the microhomology motifs we detect in our analyses of recombination breakpoints. Recombining sequences are preferentially located at short distances from each other (<100 bp). The bottom schematic shows how a chimeric read sequenced from a recombining fragment (top) would align on the source genome sequence (bottom). Dashed lines connect start and end alignment coordinates on the read to their counterparts on the genome sequence.

in vivo recombination, a prerequisite for the integration of host sequences into the DENV genome and their horizontal transfer to another host. This possibility will have to be further investigated using appropriate library preparation protocols. Importantly, a substantial amount of host RNAs can be co-encapsidated, together with the viral genome, by RNA viruses (Eckwahl *et al.*, 2016; Ghoshal *et al.*, 2015; Routh *et al.*, 2012). Whether these host RNAs, not integrated into the viral genome, can then be transferred between hosts by viruses remains an interesting question to address.

The rates of artificial recombination we detected (Table 1, column 3) are consistent with those previously reported (Di Giallonardo *et al.*, 2013; Haas *et al.*, 2011; Routh *et al.*, 2015; Shao *et al.*, 2013). They further outline the necessity to identify and remove technical chimeras from high-throughput sequencing datasets obtained by techniques using PCR, before downstream analyses (Edgar *et al.*, 2011; Routh *et al.*, 2015), if these analyses can be affected by their presence. Fortunately, the low frequency of artificial chimeric reads (<1% of all reads) and the absence of clear recombination hotspots greatly mitigate the impact of chimeras on most analyses, such as genome sequencing and assembly, variant detection and DNA conformation analyses, as these rely on the consistency between several reads obtained from a given locus or variant. The number of non-chimeric reads should largely exceed that of chimeric reads at any locus.

However, analyses that specifically search for rare recombination events will be affected by artificially-produced recombination, and we do not see a reliable criterion to determine any individual chimera as biological, unless the mechanism that created it is readily apparent (*e.g.*, transposition), or evidence indicate *in vivo* replication of the chimera (*e.g.*, Gilbert *et al.* 2016). Indeed, artificial chimeras share features with biological chimeras (Routh *et al.*, 2015): both often involve microhomology and short distances between recombining fragments. However, artificial recombination mostly joins fragments in opposite orientations (>75% of breakpoints), in stark contrast to biological recombination in FHV (~1.5%). Coupled to a bias-free library preparation protocol and an internal control to test the occurrence of *in vitro* recombination (*e.g.*, mixing DNA/RNA of species that were never in contact), this distinctive feature may be useful to globally evaluate the presence of artificial recombination in datasets obtained from short-read sequencing technology.

ACKNOWLEDGMENTS

We thank Richard Cordaux for valuable comments during the design of the study and on earlier versions of the manuscript. This work was supported by Agence Nationale de la Recherche Grant ANR-15-CE32-0011-01 TransVir (to C.G.), Agence Nationale de la Recherche Grant ANR-17-ERC2-0016-01 (to L.L.), the French Government's Investissement d'Avenir program Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases Grant ANR-10-LABX-62-IBEID (to L.L.), the City of Paris Emergence(s) program in Biomedical Research (to L.L.), the 2015–2020 State-Region Planning Contract and European Regional Development Fund, and intramural funds from the Centre National de la Recherche Scientifique and the University of Poitiers (to J.P. and C.G.).

LITERATURE CITED

Bartenschlager, R., and S. Miller, 2008 Molecular aspects of Dengue virus replication. Future Microbiol. 3(2): 155–165. https://doi.org/10.2217/17460913.3.2.155

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10(1): 421. https://doi.org/10.1186/1471-2105-10-421

Cascone, P. J., C. D. Carpenter, X. H. Li, and A. E. Simon, 1990 Recombination between satellite RNAs of turnip crinkle virus. EMBO J. 9: 1709–1715.

Clark, J. M., 1988 Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. 16(20): 9677–9686. https://doi.org/10.1093/nar/16.20.9677

Cocquet, J., A. Chong, G. Zhang, and R. A. Veitia, 2006 Reverse transcriptase template switching and false alternative transcripts. Genomics 88(1): 127–131. https://doi.org/10.1016/j.ygeno.2005.12.013

Di Giallonardo, F., O. Zagordi, Y. Duport, C. Leemann, B. Joos *et al.*, 2013 Next-Generation Sequencing of HIV-1 RNA Genomes: Determination of Error Rates and Minimizing Artificial Recombination. PLoS One 8(9): e74249. https://doi.org/10.1371/journal.pone.0074249

Eckwahl, M. J., H. Arnion, S. Kharytonchyk, T. Zang, P. D. Bieniasz *et al.*, 2016 Analysis of the human immunodeficiency virus-1 RNA packageome. RNA 22(8): 1228–1238. https://doi.org/10.1261/rna.057299.116

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight, 2011 UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16): 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Fansiri, T., A. Fontaine, L. Diancourt, V. Caro, B. Thaisomboonsuk *et al.*, 2013 Genetic Mapping of Specific Interactions between Aedes aegypti Mosquitoes and Dengue Viruses. PLoS Genet. 9(8): e1003621. https://doi.org/10.1371/journal.pgen.1003621

Feschotte, C., and C. Gilbert, 2012 Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13(4): 283–296. https://doi.org/10.1038/nrg3199

Fontaine, A., D. Jiolle, I. Moltini-Conclois, S. Lequime, and L. Lambrechts, 2016 Excretion of dengue virus RNA by Aedes aegypti allows non-destructive monitoring of viral dissemination in individual mosquitoes. Sci. Rep. 6: 24885.

Fraser, M. J., G. E. Smith, and M. D. Summers, 1983 Acquisition of Host Cell DNA Sequences by Baculoviruses: Relationship Between Host DNA Insertions and FP Mutants of Autographa californica and Galleria mellonella Nuclear Polyhedrosis Viruses. J. Virol. 47: 287–300.

García, P. B., N. L. Robledo, and A. L. Islas, 2004 Analysis of non-template-directed nucleotide addition and template switching by DNA polymerase. Biochemistry (Mosc.) 43(51): 16515–16524. https://doi.org/10.1021/bi0491853

Ghoshal, K., J. Theilmann, R. Reade, A. Maghodia, and D. Rochon, 2015 Encapsidation of Host RNAs by Cucumber Necrosis Virus Coat Protein during both Agroinfiltration and Infection. J. Virol. 89(21): 10748–10761. https://doi.org/10.1128/JVI.01466-15

Gilbert, C., and R. Cordaux, 2017 Viruses as vectors of horizontal transfer of genetic material in eukaryotes. Curr Opin Virol. 25:16–22.

Gilbert, C., J. Peccoud, A. Chateigner, B. Moumen, R. Cordaux *et al.*, 2016 Continuous Influx of Genetic Material from Host to Virus Populations. PLoS Genet. 12(2): e1005838. https://doi.org/10.1371/journal.pgen.1005838

Gire, S. K., A. Goba, K. G. Andersen, R. S. G. Sealfon, D. J. Park *et al.*, 2014 Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science 345(6202): 1369–1372. https://doi.org/10.1126/science.1259657

Goic, B., N. Vodovar, J. A. Mondotte, C. Monot, L. Frangeul *et al.*, 2013 RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. Nat. Immunol. 14(4): 396–403. https://doi.org/10.1038/ni.2542

Goic, B., K. A. Stapleford, L. Frangeul, A. J. Doucet, V. Gausson *et al.*, 2016 Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nat. Commun. 7: 12410. https://doi.org/10.1038/ncomms12410

Görzer, I., C. Guelly, S. Trajanoski, and E. Puchhammer-Stöckl, 2010 The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing. J. Virol. Methods 169(1): 248–252. https://doi.org/10.1016/j.jviromet.2010.07.040

Guieysse, A. L., D. Praseuth, M. Grigoriev, A. Harel-Bellan, and C. Hélène, 1995 Oligonucleotide-directed switching of DNA polymerases to a dead-end track. Biochemistry (Mosc.) 34(28): 9193–9199. https://doi.org/10.1021/bi00028a032

Guzman, M. G., and E. Harris, 2015 Dengue. Lancet 385(9966): 453–465. https://doi.org/10.1016/S0140-6736(14)60572-9

Haas, B. J., D. Gevers, A. M. Earl, M. Feldgarden, D. V. Ward *et al.*, 2011 Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21(3): 494–504. https://doi.org/10.1101/gr.112730.110

Jaworski, E., and A. Routh, 2018 ClickSeq: replacing fragmentation and enzymatic ligation with click-chemistry to prevent sequence chimeras. Methods Mol Biol. 1712:71–85.

Kanagawa, T., 2003 Bias and artifacts in multitemplate polymerase chain reactions (PCR). J. Biosci. Bioeng. 96(4): 317–323. https://doi.org/10.1016/S1389-1723(03)90130-7

Katzourakis, A., and R. J. Gifford, 2010 Endogenous viral elements in animal genomes. PLoS Genet. 6(11): e1001191. https://doi.org/10.1371/journal.pgen.1001191

Krupovic, M., N. Zhi, J. Li, G. Hu, E. V. Koonin *et al.*, 2015 Multiple layers of chimerism in a single-stranded DNA virus discovered by deep sequencing. Genome Biol. Evol. 7(4): 993–1001. https://doi.org/10.1093/gbe/evv034

Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods 9(4): 357–359. https://doi.org/10.1038/nmeth.1923

Lasken, R. S., and T. B. Stockwell, 2007 Mechanism of chimera formation during the Multiple Displacement Amplification reaction. BMC Biotechnol. 7(1): 19. https://doi.org/10.1186/1472-6750-7-19

Li, Y., and L. A. Ball, 1993 Nonhomologous RNA recombination during negative-strand synthesis of flock house virus RNA. J. Virol. 67: 3854–3860.

Li, D., W. B. Lott, K. Lowry, A. Jones, H. M. Thu *et al.*, 2011 Defective Interfering Viral Particles in Acute Dengue Infections. PLoS One 6(4): e19447. https://doi.org/10.1371/journal.pone.0019447

Martin, D. P., P. Biagini, P. Lefeuvre, M. Golden, P. Roumagnac *et al.*, 2011 Recombination in Eukaryotic Single Stranded DNA Viruses. Viruses 3(12): 1699–1738. https://doi.org/10.3390/v3091699

Meyerhans, A., J. P. Vartanian, and S. Wain-Hobson, 1990 DNA recombination during PCR. Nucleic Acids Res. 18(7): 1687–1691. https://doi.org/10.1093/nar/18.7.1687

Miller, D. W., and L. K. Miller, 1982 A virus mutant with an insertion of a copia-like transposable element. Nature 299(5883): 562–564. https://doi.org/10.1038/299562a0

Nagy, P. D., and J. J. Bujarski, 1996 Homologous RNA recombination in brome mosaic virus: AU-rich sequences decrease the accuracy of crossovers. J. Virol. 70: 415–426.

Nagy, P. D., and A. E. Simon, 1997 New Insights into the Mechanisms of RNA Recombination. Virology 235(1): 1–9. https://doi.org/10.1006/viro.1997.8681

Odelberg, S. J., R. B. Weiss, A. Hata, and R. White, 1995 Template-switching during DNA synthesis by Thermus aquaticus DNA polymerase I. Nucleic Acids Res. 23(11): 2049–2057. https://doi.org/10.1093/nar/23.11.2049

O'Neil, D., H. Glowatz, and M. Schlumpberger, 2013 Ribosomal RNA depletion for efficient use of RNA-seq capacity in Current Protocols in Molecular Biology, edited by Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, and J. G. Seidman *et al.* John Wiley & Sons, Inc., Hoboken, NJ.

Pääbo, S., D. M. Irwin, and A. C. Wilson, 1990 DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. 265: 4718–4721.

Patel, R., M. Lin, M. Laney, N. Kurn, S. Rose *et al.*, 1996 Formation of chimeric DNA primer extension products by template switching onto an annealed downstream oligonucleotide. Proc. Natl. Acad. Sci. USA 93(7): 2969–2974. https://doi.org/10.1073/pnas.93.7.2969

Peccoud, J., V. Loiseau, R. Cordaux, and C. Gilbert, 2017 Massive horizontal transfer of transposable elements in insects. Proc. Natl. Acad. Sci. USA 114(18): 4721–4726. https://doi.org/10.1073/pnas.1621178114

Pérez-Losada, M., M. Arenas, J. C. Galán, F. Palero, and F. González-Candelas, 2015 Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30: 296–307. https://doi.org/10.1016/j.meegid.2014.12.022

Pogany, J., J. Romero, and J. J. Bujarski, 1997 Effect of 5′ and 3′ terminal sequences, overall length, and coding capacity on the accumulation of defective RNAs associated with broad bean mottle bromovirus in planta. Virology 228(2): 236–243. https://doi.org/10.1006/viro.1996.8377

R Core Team, (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raffo, A. J., and W. O. Dawson, 1991 Construction of tobacco mosaic virus subgenomic replicons that are replicated and spread systemically in tobacco plants. Virology 184(1): 277–289. https://doi.org/10.1016/0042-6822(91)90844-2

Routh, A., and J. E. Johnson, 2014 Discovery of functional genomic motifs in viruses with ViReMa-a Virus Recombination Mapper-for analysis of next-generation sequencing data. Nucleic Acids Res. 42(2): e11. https://doi.org/10.1093/nar/gkt916

Routh, A., T. Domitrovic, and J. E. Johnson, 2012 Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. Proc Natl Acad Sci U A 109(6): 1907–1912. https://doi.org/10.1073/pnas.1116168109

Routh, A., S. R. Head, P. Ordoukhanian, and J. E. Johnson, 2015 ClickSeq: Fragmentation-Free Next-Generation Sequencing via Click Ligation of

Adaptors to Stochastically Terminated 3′-Azido cDNAs. J. Mol. Biol. 427(16): 2610–2616. https://doi.org/10.1016/j.jmb.2015.06.011

Shammas, F. V., R. Heikkilä, and A. Osland, 2001 Fluorescence-based method for measuring and determining the mechanisms of recombination in quantitative PCR. Clin. Chim. Acta Int. J. Clin. Chem. 304(1-2): 19–28. https://doi.org/10.1016/S0009-8981(00)00374-0

Shao, W., V. F. Boltz, J. E. Spindler, M. F. Kearney, F. Maldarelli *et al.*, 2013 Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA. Retrovirology 10(1): 18. https://doi.org/10.1186/1742-4690-10-18

Simon-Loriere, E., and E. C. Holmes, 2011 Why do RNA viruses recombine? Nat. Rev. Microbiol. 9(8): 617–626. https://doi.org/10.1038/nrmicro2614

Vaz, P. K., T. J. Mahony, C. A. Hartley, E. V. Fowler, N. Ficorilli *et al.*, 2016 The first genome sequence of a metatherian herpesvirus: Macropodid herpesvirus 1. BMC Genomics 17: 70.

Weiss, R. A., 2017 Exchange of Genetic Sequences Between Viruses and Hosts, pp. 1–29 in *Viruses*, *Genes*, *and Cancer*, edited by Hunter, E., and K. Bister. Springer International Publishing, Cham. https://doi.org/10.1007/82_2017_21

White, K. A., and T. J. Morris, 1995 RNA determinants of junction site selection in RNA virus recombinants and defective interfering RNAs. RNA N. Y. N 1: 1029–1040.

Zanini, F., J. Brodin, J. Albert, and R. A. Neher, 2017 Error rates, PCR recombination, and sampling depth in HIV-1 whole genome deep sequencing. Virus Res. 239: 106–114. https://doi.org/10.1016/j.virusres.2016.12.009

Zeng, X.-C., and S.-X. Wang, 2002 Evidence that BmTXK beta-BmKCT cDNA from Chinese scorpion Buthus martensii Karsch is an artifact generated in the reverse transcription process. FEBS Lett. 520(1-3): 183–184, author reply 185. https://doi.org/10.1016/S0014-5793(02)02812-0

Communicating Editor: B. Andrews