Exploratory re-encoding of Yellow Fever Virus genome: new insights for the design of live-attenuated viruses

Short title: **Exploratory re-encoding of Yellow Fever Virus genome**

R. Klitting[1*], T. Riziki[1], G. Moureau[1], G. Piorkowski[1], X. de Lamballerie[1]

[1]UMR EPV, "Émergence des Pathologies Virales", Aix-Marseille University - IRD 190 - Inserm 1207 – EHESP – IHU Méditerranée Infection), Marseille, France

*Corresponding Author

E-mail: raphaelle.klitting@posteo.de (RK)

Abstract

Virus attenuation by genome re-encoding is a pioneering approach for generating live-attenuated vaccine candidates. Its core principle is to introduce a large number of slightly deleterious synonymous mutations into the viral genome to produce a stable attenuation of the targeted virus. The large number of mutations introduced is supposed to guarantee the stability of the attenuated phenotype by lowering the risks of reversion and recombination for re-encoded sequences. In this prospect, identifying mutations with low fitness cost is pivotal as it would allow increasing the number of mutations introduced and thus, generating more stably attenuated viruses. Here, we sought to identify mutations with low deleterious impact on the *in vivo* replication and virulence of yellow fever virus (YFV). Following comparative bioinformatic analyses of flaviviral genomes, we categorized synonymous transition mutations according to their impact on CpG/UpA composition and secondary RNA structures. We designed 17 re-encoded viruses with 100-400 synonymous mutations in the NS2A-to-NS4B coding region of YFV *Asibi* and *Ap7M* (hamster-adapted) genomes. Each virus contained a panel of synonymous mutations designed according to the abovementioned categorisation criteria. Parent and re-encoded viruses were compared *in vitro* (cell culture competition experiments) and *in vivo* (hamster infection and immunisation experiments), combined to extensive genome deep-sequencing. Most of re-encoded strains showed no significant decrease in replicative fitness *in vitro*, but exhibited attenuated phenotypes *in vivo*, with alleviated virulence and in some instances, decreased replicative fitness. Overall, simple transitions that did not bring new CpG/UpA dinucleotides had the lowest deleterious effect on YFV replication and virulence *in vivo* and allowed the production of YFV strains exhibiting a mild attenuation in hamsters. In addition, the most attenuated of the re-encoded strains induced a robust, protective immunity in hamsters upon challenge with *Ap7M* virus. On this basis, we conclude that the introduction of transitions with no increase in the number of CpG/UpA dinucleotides may constitute a ground rule for the finely tuned design of massively re-encoded viral genomes.

Introduction

The *Flavivirus* genus (family *Flaviviridae)* brings together 53 taxonomically recognized species and at least 19 “tentative” species (2). This ensemble is endowed with an exceptional diversity in ecological networks that widely correlates with the phylogenetic relationships within the genus (3-6). To date, a majority of identified flaviviruses are arthropod-borne viruses (arboviruses), for which transmission amongst vertebrate hosts is ensured by mosquitoes (MBFV) or ticks (TBFV). The genus also includes viruses with no known vector (“NKV”) and some that infect only insects (insect-specific flaviviruses, “ISFV”) (7, 8). Several important (re)-emerging human pathogens fall within the MBFV and TBFV groups, notably dengue virus (DENV), yellow fever virus (YFV), tick-borne encephalitis virus (TBEV), Japanese encephalitis virus (JEV), West Nile virus (WNV) and the recently emerged, Zika virus (ZIKV). Flaviviruses consist in single-stranded positive sense RNA viruses whose virions consist in spherical, enveloped particles of *ca.* 50nm diameter. The capped, 11 kilobases genome includes a single open reading frame (ORF), flanked at its 5’ and 3’ termini by structured, non-coding regions, necessary to viral RNA translation and replication (9). ORF translation gives rise to a polyprotein that is cleaved co- and post-translationally into 3 structural (C-prM-E) and 7 non-structural proteins (NS1-2A-2B-3-4A-4B-5) (10, 11).

YFV is the founding member of the genus *Flavivirus*, that owes its name (*flavus* is the Latin word for yellow) to the jaundice associated with the liver dysfunction characteristic of yellow fever disease. In humans, YFV infects primarily the liver, causing a severe viscerotropic infection in association with high fever (12). The severity of YF infections ranges from "inapparent" cases (*i.e.* that do not call medical attention) to fatal diseases with a mortality rate between 20 and 50% amongst symptomatic cases (13). YFV natural upkeep is conditioned by the presence of both its non-human primate (NHP) host(s) and mosquito vector(s), that ensure viral maintenance through a sylvatic transmission cycle. YFV maintenance takes place in the

tropical regions of both the African and South-American continents, where human cases are primarily spill-overs from the sylvatic transmission cycle (14-18), as illustrated by the recent epidemic that occurred in Brazil in 2016-2017 (18). In some instances in Africa, human-mosquito-human transmission cycles may occur, leading to epidemics that can reach wide extents. The most recent of these large-scale African epidemics happened in Angola and Democratic Republic of Congo in 2015 and 2016 (19, 20).

YFV live-attenuated vaccine (strain 17D) was developed in 1936 by M.Theiler (21). It served as early as 1938 in Brazil (22, 23) and two of its substrains (17DD and 17D-204) are still widely used for vaccine manufacturing (24). Throughout decades of use of 17D vaccine, with 20 to 60 million doses distributed annually (25), YF vaccination has proven to be safe and efficient, providing a long-lasting immunity (26) with rare adverse events (25). The YF 17D strain was obtained through serial passages (>200) of the wild-type (WT) *Asibi* strain in mouse and chicken embryo tissues (26). Similar empirical methods allowed the production of notable live-attenuated vaccines including those against poliovirus (27), measles (28-30) and mumps (31). This strategy relies on attenuation mechanisms that involve a limited number of attenuating mutations that are, most often, non-synonymous. Such live-attenuated viruses are subject to (i) attenuation reversion, as described in the case of the poliovirus vaccine (32); (ii) generation of new biological properties, as illustrated by the gain of neurovirulence observed for the YFV French neurotropic vaccine strain (33); and (iii) recombination, as documented for the poliovirus (32, 34, 35).

A new, promising, approach for virus attenuation called codon re-encoding, has been developed a decade ago by Burns (36) and Mueller (37). Based on the initial observation that usage amongst synonymous codons is highly non-random in the genome of viruses, they hypothesised that this equilibrium could be modified in a manner that should be deleterious for virus replication. They provided the first evidence supporting this concept by changing the use of synonymous codons within the poliovirus genome, after what they could observe a decrease

in viral replication capacity. By introducing a large number of slightly deleterious mutations into the coding region(s) of the viral genome without modifying the encoded protein(s), the codon re-encoding strategy allows to bypass the limitations encountered with empirical attenuation methods: the high number of mutations involved decreases drastically the risk of reversion of the attenuating mutations and reduces the likeliness of vaccine strains recombination (with either vaccine or wild-type viruses), whilst the use of silent mutations significantly lowers the chance of emergence of undesired biological properties.

A first explanation for codon usage bias was that codon abundance correlated with that of isoaccepting tRNAs and could influence the level of protein production within a given host (38). Several other mechanisms have been proposed since then, including implication in mRNA structure and folding (39, 40), microRNA-targeting (41, 42) and enhanced recognition by the immune system (43), that may vary according to the host (44, 45). Specific and random re-encoding approaches have been efficiently applied to several human RNA viruses including poliovirus, influenza A virus, human immunodeficiency virus, respiratory syncytial virus, chikungunya virus (CHIKV), JEV, TBEV and DENV (36, 37, 46-48). Considering a major role of genome-wide mutational processes in the shaping of synonymous sites (49), specific re-encoding approaches include codon and codon-pair deoptimization as well as increase of CpG/UpA dinucleotide frequency. On the other hand, the efficiency of random codon re-encoding strategies for the attenuation of CHIKV *in vitro* and of TBEV *in vivo* suggests an important influence of local constraints (*e.g.* cis-acting sequences, miRNA targeting) on genome encoding (46, 47, 50, 51).

A lot of work lies ahead to characterize the multiple mechanisms that shape synonymous codon usage and appreciate their contribution to the attenuation process during re-encoding. In this exploratory work, we defined several types of silent mutations and sought to identify which one(s) had the least deleterious effect on viral replicative fitness. In the future, such silent changes could be introduced in large numbers along the genome of vaccine candidates.

YFV offered a suitable experimental model because (i) the virus can conveniently be produced and modified using the Infectious Subgenomic Amplicons (ISA) reverse genetics method (52) and (ii) an excellent animal model of infection exists in juvenile Syrian Golden hamsters, which would allow the comparison of wild-type and re-encoded strains *in vivo*. This model was established in our laboratory using the *Asibi*-derived, YF *Ap7M* strain (Klitting *et al.*, accepted manuscript) derived from the previously described hamster-virulent YF *Asibi*/hamster p7 strain (47). When inoculated to hamsters, both strains induce a disease with features close to that of the human disease.

Starting with a bioinformatic analysis of flaviviral genomes, we defined several types of mutations and used them for the design of 8 re-encoded sequences. The re-encoded viruses were derived from the wild-type *Asibi* strain, and the hamster-adapted *Ap7M* strain, with which they were compared both *in vitro* and *in vivo* through fitness assays, infection assays and complete genome deep-sequencing. Although only the most heavily re-encoded virus (741 mutations) showed a significant decrease in replicative fitness *in vitro*, mild to strong attenuation was observed for all the re-encoded variants *in vivo*. The lowest deleterious effect was observed for strains that were re-encoded with transitions that did not bring new CpG/UpA dinucleotides. In addition, we showed that the strains with the most attenuated phenotypes *in vivo* could induce robust protective immunity in hamsters upon challenge with the lethal YF *Ap7M* virus.

Results

In silico analysis, re-encoded strains design and production.

The results of the in silico analysis are provided in the Supplementary Results S2 section.

Framework design. Previous analyses on the impact of random codon re-encoding on chikungunya (CHIKV) and tick-borne encephalitis (TBEV) viruses, were used as a starting point to obtain observable *in vitro* and/or *in vivo* attenuation. Nougairède and colleagues observed a decrease in CHIKV replicative fitness *in vitro*, using 264 to 882 mutations located within the nsp1, nsp4 and/or envelope proteins coding regions (46). Clear *in vivo* attenuation of the TBEV was achieved by de Fabritus and colleagues using as few as 273 mutations located in the NS5 coding region (47). Based on those results, we defined a framework for the design of YFV re-encoded strains. We chose to re-encode the region located within the coding sequences of proteins NS2A to NS4B and, for each variant, used 350 synonymous mutations at the most (*ca.* 1 change every 10 nucleotides). To design the highly re-encoded variant rTs4, 388 additional mutations were introduced in the NS5 coding region. In all cases, the structural proteins coding sequence remained unchanged, which was an advantage as the production of hamster-virulent YFV variants required the use of a modified E protein sequence. To lower the risk of major deleterious effects of re-encoding (*e.g.* disruption of cis-acting sequences), we analysed an alignment of 35 YFV complete coding sequences (CDS) and defined only synonymous sites for which at least one mutation could be observed as eligible for re-encoding. Those will further be referred to as “editable sites”. Finally, we avoided creating or removing any rare codon during the re-encoding process (see definition in Supplementary Protocols S1).

Following an *in silico* analysis of flaviviral genomes, we defined two major types of synonymous

mutations (see below). The creation of additional TCG trinucleotide patterns was considered as a “specific” type of mutation. In the genus *Flavivirus*, TCG trinucleotides are frequent in Insect-Specific Flaviviruses (ISF) genomes and sparse in both No Known Vector viruses’ (NKV) and arboviruses’ genomes (see Supplementary results). In addition, this pattern may enhance the innate immunity in vertebrate cells, as suggested by Greenbaum, Jimenez-Baranda and colleagues (44, 53). “Specific” mutations also included the creation of CpG and UpA dinucleotides, which were the least frequently introduced dinucleotides (CpG and UpA) in YFV species (see Supplementary results). This is in accordance with previous observation of dinucleotide biases in the genomes of many viruses, in a fashion that was suggested to -at least partly- mirror the dinucleotide usage of their host(s) (54, 55). In vertebrates, CpG dinucleotide bias is commonly seen as the consequence of cytosine methylation-deamination mechanism (56-58) and the detection of such patterns has been proposed as a part of host-defence mechanisms towards non-self RNA (59-61). The introduction of CpG and UpA dinucleotides has been frequently used in re-encoding studies (43, 62-64). By contrast, “non-specific” types of mutations did not cause any increase in CpG/UpA dinucleotides. This category includes transitions located on and outside sites associated to the putative secondary structures identified in YFV genome during *in silico* analysis (see Supplementary results). A description of all mutation types defined in this study is available in Table 1.

	Non-specific	Specific
Mutations	-Simple transitions -Transitions on sites associated to putative secondary structures	- CpG/UpA dinucleotides transitions - TCG trinucleotides introduction

Table 1. *Mutations categories.* Based on bioinformatical analysis, 2 major types of mutations were defined.

In total, 9 re-encoded viruses were designed based on the *Asibi* strain sequence. Asibi-derived re-encoded viruses were designed with 100 to 400 mutations located between positions 3,924 and 6,759 of YF *Asibi* complete coding sequence (CDS). The heavily re-encoded strain rTs4 included 388 additional mutations located between positions 6,846 and 9,765. Finally, 8

hamster-adapted re-encoded strains were created from the *Asibi*-derived variants (other than the heavily re-encoded rTs4 virus) by introducing exactly the same mutations in the CDS of the hamster-virulent, YFV *Ap7M* virus. Importantly, all the mutations included in re-encoded strains were synonymous and were exclusively introduced at editable sites (see definition above). Posterior to re-encoding the GC% and Effective Number of Codons (ENC) were calculated for each sequence and compared to those obtained from a dataset of 35 YFV genome sequences (non-vaccinal strains) using SSE software (v1.3) (1). Apart for the most re-encoded one, rTs4 (GC%:52,3%; ENC:52,1), all the strains fell within the YFV species range in terms of both GC% (48,9%-50,5%) and ENC (52,4 - 54,7). Details regarding re-encoded strains design are available in Table 2.

Design of re-encoded strains including only "non-specific" transitions. Re-encoded strains YFV rTs1, rTs2 and rTs3, were designed by introducing increasing numbers of "simple" transitions (siTs) in the CDS of the reference strain *Asibi* (AY640589). The term "simple" refers to synonymous, non-specific (no CpG/UpA introduction) mutations located outside putative secondary structures sites. Strain YFV rTs4 included all mutations of YFV rTs3 and 388 additional non-specific transitions located between position 6,846 and 9,765 of the CDS, (741 transition in total).

Strain YFV rSS was re-encoded using 289 siTs as well as 50 non-specific transitions on sites corresponding to putative secondary structures (SII-Ts) identified in the parent sequence (described in the Supplementary Results S2).

Design of re-encoded strains including "specific" transitions. Viruses YFV rUA and rCG were designed by increasing the number of UpA and CpG dinucleotides into the viral sequence. Transitions leading to UpA dinucleotide introduction (UpA-Ts) were combined to siTs to design strain rUA (59 additional UpA). Similarly, rCG involved a combination of siTs and

CpG-Ts (59 additional CpG).

To investigate the impact of introducing TCG patterns on viral replication, mutations allowing the creation of TCG trinucleotides (TCG-Ts) were used for the design of re-encoded strain YFV rTCG (118 additional TCG). Because of the scarcity of editable sites allowing TCG introduction, design rules were relaxed: both transitions and transversions were used and some mutations affected sites corresponding to putative secondary structures (total:12) as well as non-editable sites (total:11).

Finally, a combinatory strategy was used for YFV strain rN design. It included 233 mutations (179 transitions and 54 transversions) that led to the introduction of 83 UpA and 54 CpG dinucleotides into the viral sequence.

Production of YFV re-encoded strains. All the strains used in this study (1 WT, 1 hamster-adapted and 17 re-encoded) were produced using the ISA method, that allows to recover infectious viruses after transfection into permissive cells of overlapping subgenomic DNA fragments covering the entire genome (52). All fragments combination schemes used for virus production are detailed in Fig 1 and in the “Methods” section.

Parent (*Asibi* and *Ap7M*) and re-encoded viruses were recovered after amplification and transfection of the 3 corresponding subgenomic amplicons in BHK21 cells. The viral culture supernatant was passaged twice onto Vero (*Asibi*-derived strains) or BHK21 (*Ap7M*-derived strains) cells and a third time onto BHK21 cells (all strains) before storage, next-generation sequencing and either *in vitro* or *in vivo* study.

Name	Description	Transitions	Transversions	CpG-Ts	UpA-Ts	SII-Ts	TCG-Ts	NE sites-Ts	Total	GC content (%)	ENC	Titre (TCID50/mL)
Asibi	*Asibi*	0	0	0	0	0	0	0	0	49,7	52,8	1E+08
rTs1	siTs-low	101	0	0	0	0	2	0	101	49,7	52,9	4E+05
rTs2	siTs-intermediate	247	0	0	0	0	2	0	247	49,5	52,7	4E+07
rTs3	siTs-high	353	0	0	0	0	3	0	353	49,4	52,6	4E+06
rUA	UpA-Ts	326	0	0	59	0	2	0	326	49,6	53,5	3E+06
rCG	CpG-Ts	325	0	59	0	0	19	0	325	49,7	52,8	3E+06
rSS	Ts on putative secondary structures	339	0	0	0	50	3	0	339	49,7	52,9	7E+06
rN	Combinatory strategy	179	54	59	83	0	17	0	233	49,5	52,7	4E+06
rTCG	Ts to TCG	46	72	111	0	12	118	11	118	49,4	52,6	2E+06
rTs4	siTs-very high	741	0	0	0	0	7	0	741	52,3	52,1	6E+03

Table 2. *Asibi and Ap7M-reencoded strains description.* For each YFV re-encoded strain, the number of mutations introduced within the CDS of the reference strain *Asibi* (Genbank AN: AY640589) are given both as a total and with details according to the type of mutation used. The following abbreviations were used: CpG-Ts: transition leading to CpG dinucleotide introduction, UpA-Ts: transition leading to UpA dinucleotide introduction; SII-Ts: transition on sites corresponding to putative secondary structures; NE sites-Ts: transition on sites for which no mutation has been reported in the 35 YFV CDS alignment. G+C content was calculated on viral CDS and is given in percent. For each viral CDS, the ENC was calculated using SEE software (v1.3) (1). Viral titres were determined on BHK21 cells and are given in TCID50/mL.

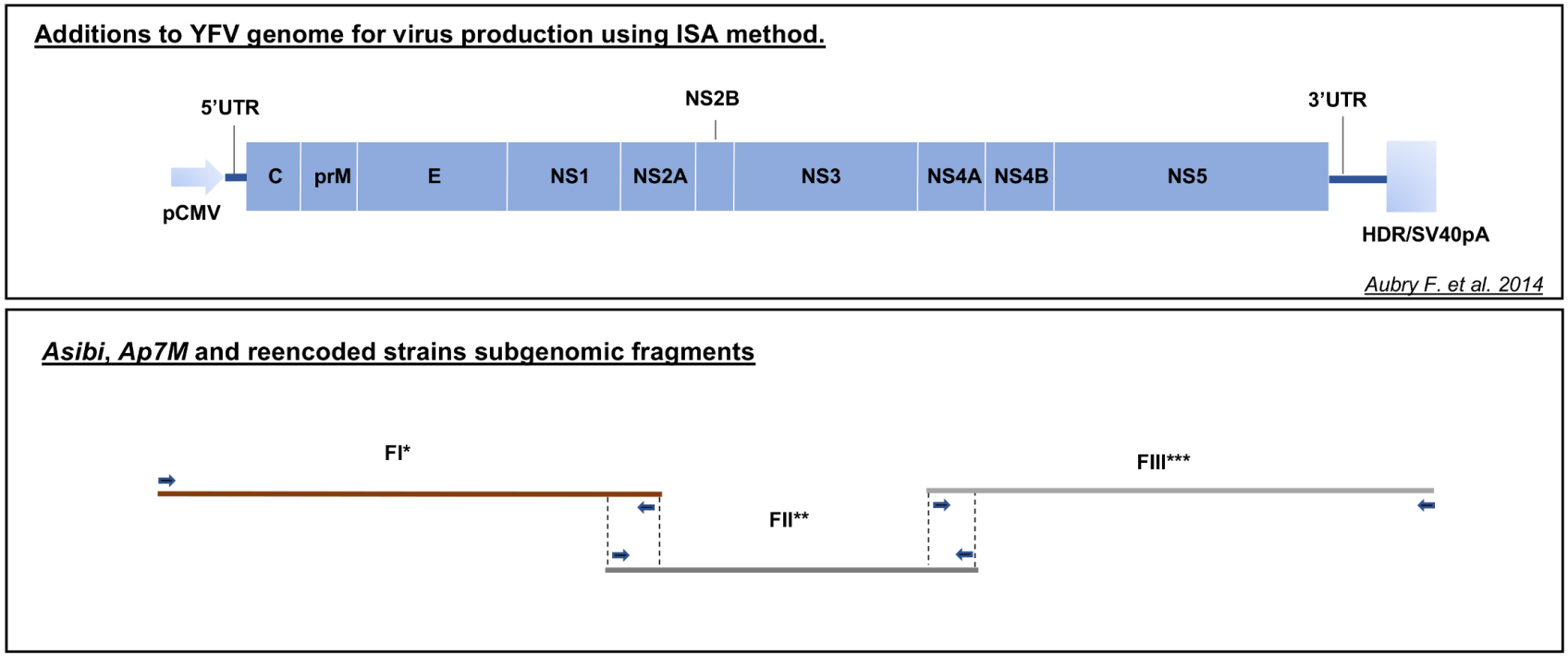

Figure 1. *Subgenomic fragments used for wild-type and reencoded virus production thanks to the Infectious Subgenomic Amplicon method.* *For *Ap7M* strain and hamster-adapted reencoded strains (Ap7M-derived) production, the subgenomic fragment FIAp7 was amplified from the corresponding plasmid and combined to subgenomic fragments FII (wild-type or reencoded) and FIII. **For each of the YFV reencoded strains (both *Asibi* and *Ap7M*-derived) production, a different subgenomic fragment FII was amplified from the corresponding plasmid. ***For the production of YFV reencoded strain rTs4, the subgenomic fragment FIIIR was amplified from the corresponding plasmid and combined to fragments FI (wild-type) and FII (reencoded).

In vitro behaviour of YFV re-encoded strains

Sequence analysis. Next Generation Sequencing (NGS) was performed on the cell culture supernatants (viral stocks *i.e.*, after 3 passages). For all viruses, both the integrity of the genome and the intra-population genetic diversity were assessed on the CDS. On the level of intra-population genetic diversity, a variant was regarded as major when the corresponding nucleotide proportion (CNP) was over 75%. Only mutations associated to major variants were considered for further analysis. All sequencing results are shown in Tables 3 and 4.

During virus production, sequence variability could arise as an artefact of the production method itself (PCR amplification of DNA fragments) and as the result of the adaptation of re-encoded strains to culture conditions.

No mutation associated to a major variant was identified in the genome of the parent strain YFV *Asibi*, which only included twelve minor variants. In re-encoded genome sequences, low numbers of mutations were observed, ranging from 0 to 6, with a mean of 2 changes/virus and 39% of NS mutations on average. Mutations were equally distributed between re-encoded and non-re-encoded regions and, nearly no convergence was observed between viral sequences. Only two viral sequences (YFV rTs2 and rCG) had mutation reversions. One was common to both strains and corresponded to a siTs mutation (C to T), at CDS position 5,430. The two others were also siTs mutation reversions (C to T and A to G, respectively), located at CDS positions 5,442 and 5,667 of YFV rTs2 and rCG, respectively. The absence of NS change in the sequence of the original *Asibi* strain and the very limited convergence between viral sequences are in accordance with the low number (3) of *in cellulo* passages. This strongly suggests that the mutations observed in viruses YFV rTs3, rCG, rSS and rTCG did not result from a selective pressure due to culture conditions but may be artefacts of the production method.

The total numbers of variants were heterogeneous, ranging from 1 to 16, with a mean of 9 variants per virus. Interestingly, variants were most frequently observed outside re-encoded regions (25% of variants in re-encoded regions), similar to what was observed for the parental sequence (8% of variants in re-encoded region). The low number of reversions and the limited sequence variability at re-encoded sites accord with previous work showing that re-encoded CHIKV strains evolved mostly through compensatory mutations rather than mutation reversions (46).

***In vitro* replicative fitness.** The infectious titre was estimated using a $TCID_{50}$ assay in BHK21 cells for both *Asibi* and *Asibi*-derived re-encoded strains using infectious cell supernatant (viral stock) (results detailed in Table 2). The maximal infectious titre was observed for the wild-type strain *Asibi* (10^8 $TCID_{50}$/mL). A 1 to 2-log decrease was observed for most re-encoded strains, with infectious titres ranging between 2.10^6 and 4.10^7 $TCID_{50}$/mL. However, the least re-encoded virus, YFV rTs1 (101 mutations), was unexpectedly endowed with a remarkably low infectious titre (4.10^5 $TCID_{50}$/mL) whilst the most re-encoded virus, YFV rTs4 (741 mutations), exhibited the lowest infectious titre (6.10^3 $TCID_{50}$/mL).

For each of the re-encoded viruses, the replicative fitness was compared to that of the wild-type strain using competition assays. Specific virus detection was achieved for one in two passages using next generation sequencing (NGS) methods by amplifying a 256 nucleotides region and counting virus-specific patterns within the read population. For all re-encoded viruses except rTs4 (≤ 353 mutations), replicative fitness *in vitro* was comparable to that of the *Asibi* strain: in all competitions, both competing viruses could still be detected at the 10th passage (patterns counts are detailed in the supplementary Table S11). By contrast, YFV the highly re-encoded rTs4 virus (741 mutations) showed a clear reduction in replicative fitness and could not be detected after the 5th passage.

Strain name	Reencoded region	Position (nt)	Protein	Mutation	Expected sequence (nt)	Observed sequence (1st variant (nt))	Observed sequence (2nd variant (nt))	1st variant proportion	2nd variant proportion
Asibi	-	-	-	-	-	-	-	-	-
rTs1	3927-6753	-	-	-	-	-	-	-	-
rTs2	3927-6753	5430	-	-	C	T	C	77,5%	22,3%
		5442	-	-	C	T	C	82,0%	18,0%
rTs3	3927-6753	4357	NS2B	M99V	A	G	A	78,3%	21,7%
		9150	-	-	A	G	A	78,5%	21,5%
rUA	3927-6753	-	-	-	-	-	-	-	-
rCG	3924-6750	4396	NS2B	F112L	T	C	T	88,2%	11,8%
		5430	-	-	C	T	-	>95%	-
		5667	-	-	A	G	-	>95%	-
		6082	NS3	C544R	T	C	-	>95%	-
rSS	3939-6759	561	-	-	T	C	-	>95%	-
		3190	NS1	I286V	A	G	-	>95%	-
		4308	-	-	G	T	-	>95%	-
		4623	-	-	A	G	-	>95%	-
		6012	-	-	A	G	-	>95%	-
rN	3924-6750	-	-	-	-	-	-	-	-
rTCG	3939-6759	2468	NS1	I44T	T	C	-	>95%	-
		2795	NS1	Q154R	A	G	-	>95%	-
		7629	-	-	T	C	-	>95%	-
		8565	-	-	T	C	-	>95%	-
		8916	-	-	T	C	-	>95%	-
		9297	-	-	T	C	-	>95%	-
rTs4	3927-9765	-	-	-	-	-	-	-	-

Table 3. *Sequence analysis of Asibi and Asibi-derived reencoded viruses.* For each virus, the changes found in the CDS from the viral stock are detailed above. The expected sequence refers to that of plasmids that served for subgenomic fragment amplification while producing viruses thanks to the ISA method. Only mutations associated to major variants (CNP>75%) are described. Amino Acid (AA) positions are given with reference to the beginning of mature protein sequence (ie. AA position 107 in prM, written prM 107, corresponds to AA position 228 within the precursor polyprotein). A detailed map of protein positions within the YFV polyprotein sequence is provided in the supplementary Table S12.

VIRUS	Mutations	Non-Synonymous mutations	Variants	Reversions	Mutations in reencoded region	Variants in reencoded region
Asibi	0	0	12	0	0	1
rTs1	0	0	12	0	0	0
rTs2	2	0	9	2	2	3
rTs3	2	1	6	0	1	2
rUA	0	0	16	0	0	5
rCG	4	2	5	2	4	1
rSS	5	1	1	0	3	0
rN	0	0	11	0	0	2
rTCG	6	2	4	0	0	1
rTs4	0	0	15	0	0	9
rTs1h	0	0	13	0	0	4
rTs2h	0	0	22	0	0	6
rTs3h	3	3	13	0	1	3
rUAh	0	0	17	0	0	5
rCGh	0	0	10	0	0	0
rSSh	0	0	19	0	0	2
rNh	5	1	4	0	0	0
rTCGh	2	2	7	0	0	0

Table 4. *Sequence analysis of Asibi,* and both *Asibi and Ap7M-derived reencoded viruses.* Mutations and variants counts are given as raw numbers. Reversion mutations were defined with reference to YFV strain *Asibi*.

In vivo behaviour of YFV re-encoded strains

The infectious titre of viral stocks for *Ap7M* and *Ap7M*-derived re-encoded strains were established in BHK21 cells, using a $TCID_{50}$ assay (results detailed in Table 5). For the majority of re-encoded viruses, no significant difference was observed with the parent strain *Ap7M* (9.10^8 TCID50/mL), whith infectious titres ranging from 4.10^8 to 6.10^8 TCID50/mL. However, viruses YFV rNh and rTCGh showed a 1 to 2-log reduction in viral titre (6.10^6 and 6.10^7 TCID50/mL, respectively).

Sequence analysis. NGS was performed on the viral stocks for all *Ap7M*-derived re-encoded viruses as described in the previous section (results shown in Tables 4 and 6).

As observed for *Asibi*-re-encoded viruses, the numbers of mutations associated to major variants were low, ranging from 0 to 5, with a mean number of 1 change/virus and 59% of non-synonymous mutations on average. Mutations were mainly distributed outside re-encoded regions (90%) with no reversion. As observed for *Asibi*, the parent *Ap7M* strain sequence did not arbour any NS change associated to major variants. Besides, the re-encoded viruses went through a limited number of *in cellulo* passages and their sequences did not exhibit any evidence of convergent evolution. For these reasons, the mutations observed in YFV rTs3h, rNh and rTCGh may be artefacts of the production method rather than the result of the adaptation of re-encoded viruses to culture conditions. Notably, for YFV rTs3h, the NS change observed in the inoculum was obviously dispensable *in vivo* as it was not maintained upon infection in hamsters.

The total number of variants per virus varied among strains, ranging from 4 to 22, with a mean number of 13 variants per virus. Similarly to what was observed for *Asibi*-derived re-encoded strains, the diversity at re-encoded sites was low, with only a minority of variants falling within the re- encoded regions (19%).

Mutant name	Titre (TCID50/mL)	Mortality	Mean viral loads (RNA copies/g of liver)			Mean normalised percentage of initial weight (%)	
			Day 3 pi	Day 6 pi	Endpoint	Day 3 pi	Day 6 pi
Ap7M	*9E+08*	88%	2E+09	9E+08	3E+08	97%	71%
rTs1h	*6E+08*	0%	7E+08^{EP}	**5E+07^{A}**	3E+07	106%	92%
rTs2h	*6E+08*	17%	1E+10	6E+08	3E+09	107%	78%
rTs3h	*6E+08*	20%	9E+08	9E+08	2E+09	104%	78%
rUAh	*4E+08*	0%	3E+09^{EP}	2E+09	**3E+07^{A}**	96%	78%
rCGh	*4E+08*	0%	7E+08^{EP}	2E+09^{EP}	**5E+06^{A}**	91%	75%
rSSh	*4E+08*	0%	1E+10	2E+09^{EP}	9E+07	89%	62%
rNh	*6E+06*	0%	1E+09^{EP}	**8E+07^{A}**	**1E+07^{A}**	97%	96%
rTCGh	*6E+07*	0%	4E+09	5E+08^{EP}	6E+07	97%	80%

Table 5. *In vivo and in vivo phenotypes of Ap7M and Ap7M-derived reencoded viruses. In vitro* infectious titres, mortality rates, mean viral loads at 3, 6 dpi and at endpoint as well as mean normalised percentage of initial weight means are given for all viruses. Viral loads, normalised percentage of initial weight and infectious titre evaluations are detailed in the corresponding paragraphs within the Materials and Methods section. Significant difference between viral loads at 3, 6 dpi and viral loads at endpoint (wilcoxon rank-sum test p-value<0.05) is indicated with a “EP” superscript next to the corresponding viral loads’ mean. Significant viral loads difference observed between *Ap7M* and other viruses at 3, 6 dpi or in endpoint (wilcoxon rank-sum test p-value<0.05) is indicated using both bold characters and an “A” superscript next to the corresponding viral loads’ mean.

***In vivo* phenotype: comparative study of pathogenicity.** Groups of 12 three-week-old female hamsters were inoculated intra-peritoneally with 5.10^5 $TCID_{50}$ of virus (either *Ap7M* or *Ap7M*-re-encoded strains) and a control group of 2 uninfected hamsters was maintained for weight monitoring. Clinical follow-up included (i) clinical manifestations of the disease (brittle fur, dehydration, prostration and lethargy), (ii) body weight (weight evolution was expressed as a normalised percentage of the initial weight (%IW)) and (iii) death. Respectively 3 and 4 hamsters were euthanised at days 3 and 6 post-infection (dpi) to conduct virology investigations from liver samples, whilst the 5 others were kept for evaluating mortality rate (endpoint in case of survival: 16 dpi). Virology follow-up was achieved by performing qRT-PCR and next generation sequencing on RNA extracted from the liver homogenates. RNA samples were subsequently pooled to obtain one sequence for each virus, reflecting genetic diversity amongst the 5 hamsters kept until death/euthanasia. For all viruses, the *in vivo* phenotype and the sequencing results are detailed in Tables 4, 5 and 6.

Nearly all hamsters inoculated with *Ap7M* strain developed outward signs of illness such as brittle fur, prostration, dehydration and lethargy. One hamster did not show any sign of illness, both its liver and blood were tested negative for YFV genome and it was excluded from analysis. A high mortality rate (88%) was observed, similar to previous observations (Klitting *et al.*, accepted manuscript). Clinical signs of the disease appeared as early as 4 dpi and all animals died within 2/3 days after onset of the symptoms. Weight loss was not observed at 3 dpi (mean of %IW: 97%), but at 6 dpi, severe weight loss was recorded (mean of %IW: 71%). All livers were found to be YFV-positive by qRT-PCR, with viral RNA loads ranging between 6.10^7 and 6.10^9 RNA copies per gram of liver and no significant difference between the viral yields at 3 and 6 dpi and at endpoint. No change was detected in the viral sequence after propagation in hamsters.

Strain name	Position (nt)	Protein	Mutation	Expected sequence (nt)	Observed sequence (consensus (nt))	AA change	Observed sequence (Inoculum (nt))		Observed sequence (Hamster (nt)		
							1st variant	2nd variant	Retrieved sequence	1st variant	2nd variant
rTs1h	836	prM	A158V	C	T	A → V	C	-	1-4806, 5430-7566	T	-
	4574	NS3	G41A	G	C	G → A	G	-		C	g
	5442	-	-	C	T	-	C	-		T	-
rTs2h	-	-	-	-	-	-	-	-	1-3858, 4653-5771, 6083-10236	-	-
rTs3h	1649	Env	T265I	C	T	T → I	C	t	1-10236	T	-
	4604	NS3	R51M	T	G	R → M	G	t		T	g
rUAh	-	-	-	-	-	-	-	-	1-6210, 6492-7565	-	-
rCGh	283	-	-	T	C	-	T	-	1-4652, 5433-7564	C	t
	1051	Env	T66A	A	G	T → A	A	-		G	a
	2541	-	-	A	G	-	A	-		G	-
rSSh	5439	-	-	A	G	-	A	-	14-10228	G	-
rNh	818	prM	I152T	T	C	I → T	C	-	7-5236, 5422-10229	C	-
	1686	-	-	T	C	-	C	-		C	-
	2052	-	-	G	A	-	A	-		A	-
	7490	NS4B	N241S	A	G	N → S	g	a		G	-
	8253	-	-	G	A	-	A	g		A	-
	9783	-	-	A	G	-	G	a		G	-
rTCGh	2222	Env	I456T	T	C	I → T	C	t	11-10223	C	t
	7107	NS4B	I113M	A	G	I → M	G	a		G	a

Table 6 *Sequence and intra-population diversity in viral culture supernatant and hamster liver samples for Ap7M and Ap7M-derived reencoded strains.* For each virus, the changes found either in the complete coding sequence from the viral stock or in partial sequences from infected hamster liver homogenates are detailed above. The expected sequence refers to that of plasmids that served for subgenomic fragment amplification while producing viruses thanks to the ISA method. For 1st and 2nd variant descriptions, letters written in capital letters correspond to variant proportions above 75%, lowercase letters to variant proportions between 6 and 75% and dashes to variant proportions below 5%. Amino Acid (AA) positions are given with reference to the beginning of mature protein sequence (ie. AA position 107 in prM, written prM 107, corresponds to AA position 228 within the precursor polyprotein). A detailed map of protein positions within the YFV polyprotein sequence is provided in the supplementary Table S12.

397 We observed a reduction in mortality rate in all groups infected with re-encoded viruses. For
398 most of them (YFV rTs1h, rUAh, rCGh, rSSh, rNh and rTCGh), lethality had completely
399 disappeared. However, for YFV rTs2h and rTs3h, mortality was drastically reduced but still
400 observed (17 and 20%, respectively).
401
402 Whatever the observed mortality rate, clinical signs of illness were observed with varying
403 degrees of severity amongst groups. No association was observed between mortality and
404 weight evolution at 3 or 6 dpi, with non-lethal viruses inducing limited (rNh, rTs1h) as well as
405 severe weight loss (rCGh, rSSh) in infected hamsters. At 3 dpi, weight evolution was
406 heterogeneous, with an increase up to 7% in YFV rTs1h, rTs2h and rTs3h-infected groups, a
407 slight decrease (between 3 and 4%) in groups infected with YFV rUAh, rNh and rTCGh, and a
408 greater decrease up to 11% in YFV rCGh and rSSh-infected groups. At 6 dpi, important weight
409 loss was observed in most groups (up to 38%) except rTs1h and rNh-infected groups (8 and
410 4%, respectively).
411
412 The evolution of viral yields at 3 and 6 dpi was not associated to mortality rates. At 3 dpi, viral
413 loads in the liver were close to values observed with YF *Ap7M*, ranging from 7.10^{8} to 1.10^{10}
414 RNA copies/gram of liver. At 6 dpi, most of them remained close to what was observed with
415 *Ap7M*. However, in some cases (YFV rTs1h and rNh), significantly lower viral loads were
416 observed (5 and 8.10^{7}, respectively). For most viruses, endpoint viral loads were close to that
417 of *Ap7M* but for YFV rUAh, rCGh and rNh, they were significantly lower (between 5.10^{6} and
418 3.10^{7}). Unsurprisingly, the only association we observed was between survival and endpoint
419 viral loads, with significantly lower viral yields in the liver of surviving hamsters.
420
421 Several elements can explain the apparent absence of association between mortality and the
422 evolution of viral loads. First, significant reductions in viral loads were always associated to a
423 complete loss of lethality (for viruses rNh and rTs1h), showing that these two criteria may be
424 at least loosely associated. Furthermore, in terms of replicative fitness and thus, of viral loads,

the threshold that corresponds the loss of lethality may be subtle and undetectable under the experimental conditions as set up here. In addition, the viral loads that are reported here are evaluated from liver samples and may be less sharply associated to the disease evolution than viral loads in the blood. Finally, during YFV infection, it is likely that the pathogenesis involves an immunological component (22, 65, 66). Hence, levels of pro- and anti- inflammatory cytokines may be more closely associated with mortality than viral loads and weight evolution. Such criteria were not included in our experimental design.

Evaluation of the phenotypic impact of re-encoding. Some conclusions can be drawn by analysing the *in vivo* phenotypes of re-encoded viruses in the light of the sequencing data obtained from infected hamsters liver samples (see details in Table 6). To begin, the sequences obtained from the liver samples of hamsters infected with the model strain *Ap7M* did not include any mutation. This indicates that the sequence of the parent virus is stable under the experimental conditions and that it does not need adaptation mutations to efficiently replicate *in vivo*.

Impact of "non-specific" re-encoding strategies. We observed a low deleterious effect of the introduction of simple transitions (siTs) (*i.e.*, with no modification of putative secondary structures) on the *in vivo* phenotype of YFV. Strain rTs2h (247 siTs) exhibited a mild attenuation *in vivo* with a decreased mortality rate, a reduction in weight loss at 6 dpi but no significant change in viral loads. As the viral sequence remained unchanged along the experiments, the *in vivo* phenotype can be rightfully interpreted as the result of genome re-encoding. rTs3h (353 siTs) and rTs1h (101 siTs) viruses showed either mild or strong attenuation *in vivo*, respectively. For both, NS mutations associated to major variants arose during the infection *in vivo* and likely resulted from an adaptation of the re-encoded viruses to *in vivo* conditions during the infection. However, they did not allow the recovery an *in vivo* phenotype similar to that of the parent strain *Ap7M*. Surprisingly, the least of the re-encoded strains, rTs1h (101 siTs) showed to be more attenuated *in vivo* than the rTs2h strain, that

includes all rTs1h mutations as well as 146 additional siTs. Similarly, we observed a mild deleterious effect of the introduction of transitions at sites involved in putative secondary structures. The strain rSSh sequence involved siTs as well as transitions located on putative secondary structures and remained stable throughout the *in vivo* experiments. The variant showed a partially attenuated phenotype *in vivo* (complete loss of mortality, no alleviation in weight loss nor viral loads) that can thus be attributed to this second re-encoding strategy. Overall, these results show the efficiency of the use of “non-specific” transitions in providing a stable re-encoded infectious genome exhibiting in vivo attenuation.

Impact of “specific” re-encoding strategies. Variants with increased CpG and UpA dinucleotides exhibited attenuated phenotypes *in vivo*, with no mortality, alleviated weight losses and no reduction in viral loads. For variant rUAh (59 UpA), the sequence remained similar to the design throughout the *in vivo* experiments. Hence, the attenuation resulted from sequence re-encoding. The inoculum sequence of variant rCGh (59 CpG) was similar to the design and NS changes associated to major variants arose during the infection *in vivo*. Those mutations are likely to be adaptive but did not result in a complete recovery of virulence *in vivo*. Altogether, these results confirm the attenuating effect of the introduction of CpG/UpA dinucleotides into viral genomes as reported on previous occasions (43, 62-64).

The variant rTCGh (118 TCG) had an inoculum sequence that differed from the original design. To conclude on the compensatory role of the NS mutations that arose during strain production, further experiments would be required. Nevertheless, the attenuated phenotype (loss of mortality, alleviation of weight loss) of the re-encoded virus upon infection in hamsters suggest a deleterious effect of TCG trinucleotides’ introduction on YFV virulence *in vivo* that was not completely outweighted by the presence of additional NS changes. Hence, this suggests that the use of TCG introduction may impact the *in vivo* phenotype of YFV upon infection in vertebrate hosts.

481 Finally, for strain rNh, NS mutations were identified in the sequences from both the inoculum
482 and hamster liver samples. It is therefore complicated to determine if those changes were
483 production artefacts or resulted from an adaptation of the virus to *in vitro* or *in vivo* conditions.
484 Still, the virus exhibited a strongly attenuated phenotype *in vivo* (no mortality, alleviated weight
485 loss and viral loads) indicating that, regardless of the emergence of the -potentially
486 compensatory- NS mutations, the combination of several re-encoding strategies allowed to
487 reach an important degree of attenuation considering the low number of mutations used for re-
488 encoding (233).
489
490 **Adaptation of re-encoded viruses through NS mutations.** Altogether, these results
491 indicate that mutations can arise and stably establish in re-encoded viruses in response to *in*
492 *vivo* conditions. This was observed for a majority -but not all- of the re-encoded viruses, notably
493 some with as few as 101 re-encoded sites (rTs1h) and as much as 353 re-encoded sites
494 (rTs3h). In most cases, those mutations did not allow a full recovery of *in vivo* viral replicative
495 fitness. Observable changes in the biological properties of the viruses (*e.g.* change in tropism,
496 as reported for YF strain 17D), were not reported in any of the viruses that exhibited NS
497 mutations.
498
499 **Immunisation potential of YF-re-encoded strains: serological investigations.** For
500 each of the re-encoded viruses that showed complete loss of mortality *in vivo*, sera were
501 recovered from infected hamsters at 16 dpi during the comparative study of pathogenicity (see
502 above). Sera from both uninfected hamsters and vaccinated humans were included as
503 negative and positive controls, respectively. Sero-neutralisation tests were performed with all
504 sera, using the YF *Ap7M* virus and BHK21 cells. Neutralising antibodies were detected in both
505 hamsters infected with YF-re-encoded viruses (IC_{50} ranging between 2.10^{-3} and 1.10^{-5}) and in
506 a control group of 17D-vaccinated humans (IC_{50} between 4 and 5.10^{-4}) whilst the negative
507 control group did not show any neutralisation activity against *Ap7M* (see Fig 2).

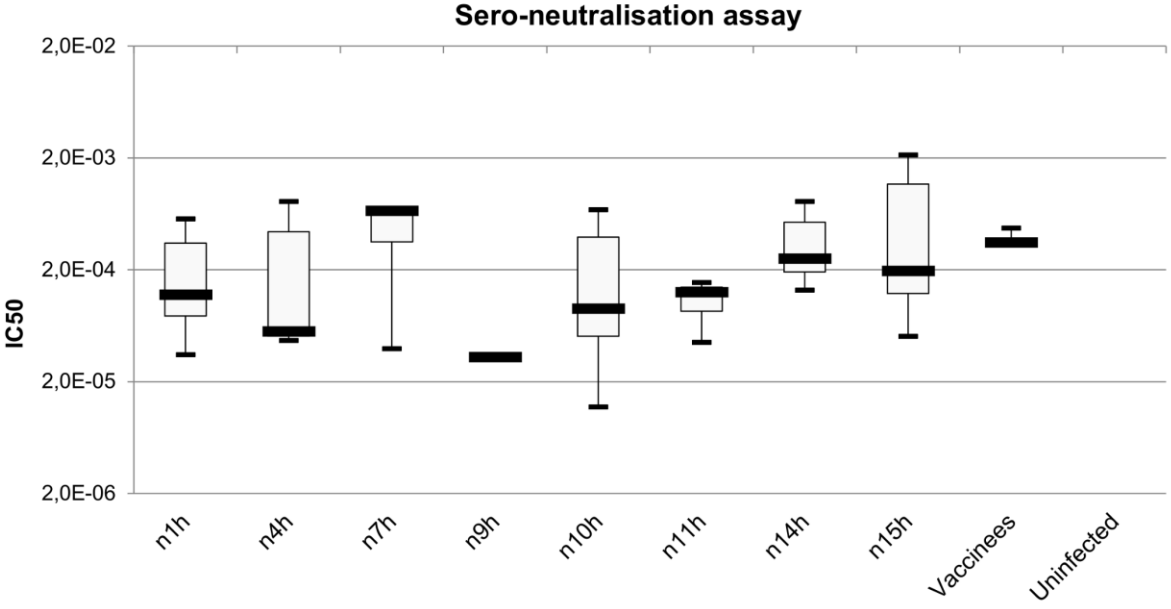

Figure 2. *Results of YFV serology at 16 days post-infection.* Sero-neutralisation as-says were performed for all re-encoded virus that showed complete loss of mortality in vivo. Sera from both uninfected hamsters and vaccinated humans (vaccinees) were used as nega-tive and positive controls. Results are given as IC50 values evaluated using Graphpad Prism Software (v7.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com). Thick black lines indicate median values and minimum/maximum val-ues are indicated by dashes.

Immunisation potential of YF-re-encoded strains: challenge experiments. Two groups of 11 three-week-old female hamsters were inoculated intra-peritoneally with 5.10^5 $TCID_{50}$ of virus (either YFV rCGh or rNh). A group of 9 hamsters was not immunised before the challenge and a control group of 2 uninfected hamsters was maintained for weight monitoring.

Unexpectedly, one hamster from the YFV rCGh infected group died at 6 dpi, although it did not show any symptom nor weight loss. Its liver was tested positive for YFV. Twenty-four days after inoculation, 2 and 3 hamsters from rCGh and rNh groups, respectively, and a hamster from the non-infected group were euthanised and both serum and liver samples were recovered for sero-neutralisation assay and viral loads determination. For all 3 groups, the 8 remaining hamsters were inoculated intra-peritoneally with 5.10^5 $TCID_{50}$ of virus (*Ap7M*). Two hamsters were euthanised at 3 and 6 days post-challenge (dpc) to conduct virology investigations from liver samples whilst the 4 others were kept for evaluating mortality rate (endpoint in case of

survival: 12 dpc). The protection was evaluated by determining for each group (i) viral loads in liver at 3 and 6 dpc and at endpoint, (ii) weight evolution at 6 and 10 dpc, (iii) survival. With regards to the number of samples tested for virology and immunology investigations, no statistical test was achieved in this section. Detailed results for this experiment are available in Table 7.

No mortality was observed post-challenge in either immunised or control groups, in coherence with previous observations indicating that mortality rate in hamsters decreases importantly in adult (>6-week-old) individuals (9). However, there was a great difference in the evolution of viral loads, with no increase in viral yields at 3 and 6 dpc in rCGh and rNh immunised animals while high viral loads were detected in the control animals, ranging around 10^9 RNA copies/g of liver. Evidence of protection was also provided by weight evolution, with no weight loss in both immunised groups and a weight loss up to 18% at 10 dpc in the control group. Altogether, those results indicate an efficient immunisation of hamsters inoculated with YFV re-encoded strains rCGh and rNh.

Mutant name	**Mortality**	**Mean viral loads (RNA copies/g of liver)**				**Mean normalised percentage of initial weight (%)**		
		Challenge	Day 3 pc	Day 6 pc	Endpoint	Challenge	Day 6 pc	Day 10 pc
Uninfected	0%	0E+00	4E+09	2E+09	6E+07	111%	83%	82%
rCGh	0%	2E+07	1E+07	3E+07	1E+07	93%	100%	98%
rNh	0%	2E+07	2E+07	2E+07	3E+07	99%	106%	110%

Table 7. *Challenge assay.* Mean viral loads at day of challenge, 3/6 dpc and endpoint well as mean normalised percentage of initial weight means at day of challenge and 3/6 dpc are given for all groups (*i.e.* immunised and control). Viral loads and normalised percentage of initial weight evaluations are detailed in the corresponding paragraphs within the Materials and Methods section.

Discussion

Genome re-encoding through synonymous codons replacement has allowed attenuating numerous RNA viruses among which poliovirus, influenza, respiratory syncytial, chikungunya, tick-borne encephalitis and dengue virus (36, 37, 46-48). Additional insights into the mechanisms contributing to the attenuation phenomenon following synonymous codons replacement should prove valuable for the design of attenuated viruses with fine-tuned phenotypes, some of which may represent relevant vaccine candidates. The initial purpose of this work was to identify types of synonymous mutations endowed with the most limited deleterious effect on viral replicative fitness. Previous codon re-encoding work was achieved in our laboratory for two other arboviruses, CHIKV and TBEV (46, 47). In the light of this experience, we built an exploratory study of the effects of re-encoding on both the *in vitro* and *in vivo* phenotypes of yellow fever virus (YFV). Our choice was motivated by the availability of both a suitable reverse genetics system for YFV re-encoded strains production (ISA method (52)) and a robust YF hamster model (67). Hence, we could test first the *in vitro*, and then the *in vivo* replicative fitness of our strains before using some of the live-attenuated candidates for hamster immunisation.

Here, we demonstrated the direct impact of synonymous substitutions on YFV replicative phenotype *in vitro* and *in vivo*. As previously described with TBEV (47), the level of re-encoding necessary to observe viral attenuation *in vivo* was significantly lower than that required to obtain an observable decrease in replicative fitness *in vitro*. All re-encoded strains with only one re-encoded fragment (101-350 mutations) maintained a replicative phenotype close to that of the parental *Asibi* strain *in vitro*. Only YFV rTs4, that included an additional re-encoded fragment (741 mutations in total), showed a severe reduction in replicative fitness *in vitro*. By contrast, *in vivo*, we reported a broad range of attenuation degrees amongst all the strains that were tested in hamsters (101-350 mutations). Whilst the least attenuated viruses showed no

significant loss of replicative fitness and retained a degree of lethality, the most attenuated ones displayed a complete loss of lethality, along with a 1 log reduction in viral loads in the liver and a significant alleviation in weight loss. We did not produce nor test a hamster-adapted version of the rTs4 strain because such a heavily re-encoded virus would have been unlikely to achieve infection *in vivo*.

In contrast with previously published re-encoding studies, deep sequencing information was made pivotal in our analysis process. Notably, we studied sequences obtained from infected hamster liver homogenates to accurately associate the viral sequence(s) to the observed phenotype. Our results indicate that, in re-encoding studies, genetic information and in particular, that retrieved from *in vivo* experiments, is critical to properly determine the significance of synonymous substitutions within the attenuation mechanism. When monitoring the evolution of YFV re-encoded strains after production (*i.e.*, transfection followed by 3 passages *in vitro*) or after one passage in hamster, we observed the emergence of both synonymous and non-synonymous substitutions. Those changes may be artefact of the production method (PCR amplification of DNA fragments, stochastic evolution events) or result from the adaptation of re-encoded viruses to *in vitro* or *in vivo* conditions of replication. In some cases, we observed the emergence of non-synonymous mutations present in major variants of the viral stock used as an inoculum for *in vivo* experiments. For those viruses, the confrontation of the results from *in vivo* experiments with sequencing data was thus limited as NS mutations impeded the identification of the specific impact of silent mutations on the viral phenotype *in vivo*.

Altogether, these results allow to draw several lines of conclusion regarding the *in vivo* phenotypic impact of the different strategies used to produce re-encoded YFV strains.

First, we observed a low deleterious effect of the introduction of simple transitions (*i.e.*, with no modification of putative secondary structures) on the replicative fitness of YFV. These simple transitions did not cause any significant modification of GC% or ENC, indicating that additional,

subtle mechanisms are at stake in the attenuation process. Similarly, we observed a mild deleterious effect of the introduction of transitions at sites involved in putative secondary structures. Contrary to secondary structures that have been identified in the untranslated regions of YFV (9), these putative structures within the virus open reading frame cannot be key to viral replication. Rather, role(s) in finely tuning the efficiency of translation (mRNA or co-translational folding) or in interactions with viral, and non-viral proteins have been suggested (39, 40, 68) .

Previous studies have showed that the increase in CpG/UpA dinucleotides is an efficient attenuation mechanism (43, 62-64). We further confirmed the deleterious effect of the introduction of CpG/UpA dinucleotides on virulence *in vivo*. We also observed that the introduction of TCG trinucleotides was affecting the *in vivo* phenotype of YFV. *In silico* analysis show that such patterns are likely to be selected against in flaviviruses infecting vertebrates. In addition, studies on influenza virus brought evidence that CpG dinucleotides in an A/U context may enhance virus detection by the host innate immune system (44, 45, 53). Such trinucleotides can be used to induce a host-dependent attenuation, as recently done for Dengue virus (69, 70).

The initial aim of this exploratory work was to identify mutation type(s) with limited deleterious effect on viral replicative fitness, usable for large scale genome re-encoding. In this regard, “non-specific” transitions (*i.e.* with no CpG/UpA creation) represent the mutation type that best met our initial criteria. On this basis, we conclude that the introduction of transitions with no CpG/UpA increase may constitute a ground rule for the finely tuned design of massively re-encoded viral genomes. This does not rule out the possible use of mutations that increase the rate of CpG/UpA dinucleotides. We propose that using a limited number of such mutations may secure the innocuity of re-encoded genomes *in vivo*, because these patterns ensure an enhanced activation of the host innate immunity. In this way, procedures combining large, pan-genomic non-specific re-encoding with limited bespoke CpG/UpA re-encoding should allow to obtain safe live-attenuated viruses with finely tuned phenotypes.

Besides deciphering the mechanisms of attenuation, evaluating the stability of viral re-encoded variants is a crucial objective for the development of vaccine candidates. Here, we did not observe any reversion from YFV re-encoded strains sequence to the parental *Ap7M* sequence *in vivo* and nearly no reversion *in vitro*. However, for several re-encoded strains, the occurrence of non-synonymous mutations in the sequences from the inoculum and/or hamster livers suggests that YFV re-encoded viruses may adapt and evolve both *in vitro* and *in vivo*. The molecular evolution of viral re-encoded strains has already been described for HIV and CHIKV-derived strains *in vitro*. The occurrence of synonymous and non-synonymous mutations was associated to partial (CHIKV) or full (HIV) WT replication capacity recovery (46, 71). In accordance with the observations made on CHIKV evolution *in vitro*, some re-encoded strains retained a strongly attenuated phenotype that was not overweighted by the emergence of adaptative mutations. Altogether, these elements suggest that, in some cases, the accumulation of slightly deleterious mutations can trap the virus into a fitness valley, with no possible mutational pathway allowing a complete restoration of its replicative fitness.

As illustrated by the case of poliovirus vaccines, live-attenuated vaccine strains have the potential to evolve upon inoculation and in case of dissemination following vaccination. Hence, the development of re-encoded vaccine candidates should go along with obtaining robust information regarding the fate of the variants *in vivo*. However, to the best of our knowledge, the evolutionary behaviour of re-encoded viruses *in vivo* has never been studied. Therefore, this represents a major step for future studies that will require relevant experimental design. Live attenuated YFV re-encoded vaccine candidates would provide a convenient experimental model by dint of the availability of appropriate reverse genetics and animal propagation systems. In this work, we mostly used YF variants re-encoded in one specific region of the genome (NS2A-NS4B), which were well adapted to our experimental objectives but would not be properly suited for evolution studies. An optimized and more realistic design would rely on the evenly distribution of synonymous mutations with low fitness weight along the genome asso-

ciated to *in vivo* serial passage and deep sequencing characterization of genetic drifting variants. Their study may provide major information regarding the stability of strains with synonymously modified codon compositions and in particular, their potential to evolve towards modified phenotypes *in vivo*.

Methods

Detailed protocols regarding in silico analysis, virus and cell culture (viral production, transfection and titration (TCID50); competition and sero-neutralisation assays), PCR amplification (subgenomic amplicon production), quantitative real-time PCR (qRT-PCR) assays, sample collection and Next-Generation Sequencing (NGS) stand in the supplementary Protocol.

Cells and animals

Viruses were produced in Baby hamster kidney BHK21 (ATCC, number CCL10) and Vero (ATCC, CCL81) cells and titrated in BHK21 cells. *In vivo* infection was performed in three-week-old female Syrian Golden hamsters (*Mesocricetus Auratus*, Janvier and Charles River laboratories).

Ethics statement

Animal protocols were reviewed and approved by the ethics committee “Comité d’éthique en expérimentation animale de Marseille—C2EA—14” (protocol number 2015042015136843-V3 #528). All animal experiments were performed in compliance with French national guidelines and in accordance with the European legislation covering the use of animals for scientific purposes (Directive 210/63/EU).

Recovery of infectious viruses and stock production

Fragments combinations. For *Asibi* strain production, 3 subgenomic fragments were used: FI (positions 1 to 3,919 in CDS and the 5’UTR (119 nucleotides (nt)), FII (3,843 to 6,838) and FIII (6,783 to 10,236 and 3’UTR (509 nt). They correspond to the complete genome of the

strain, flanked respectively at 5' and 3' termini by the human cytomegalovirus promoter (pCMV) and the hepatitis delta ribozyme followed by the simian virus 40 polyadenylation signal (HDR/SV40pA) (pCMV and HDR/SV40pA sequences were described by Aubry and colleagues (52). The hamster-virulent strain YFV *Ap7M*, was obtained from *Asibi* virus by substituting the FI subgenomic fragment by the FIAp7 fragment, that includes 10 hamster-virulence mutations as described elsewhere (Klitting *et al.*, accepted manuscript).

Re-encoded strains rTs1, rTs2, rTs3, rUA, rCG, rSS, rN and rTCG were derived from *Asibi* by combining a re-encoded version of the 2nd subgenomic fragment FII to the WT 1st (FI) and 3rd (FIII) subgenomic fragments. For the heavily re-encoded rTs4 strain, the WT fragment FI was combined to the re-encoded FII fragment of strain rTs3 and to a re-encoded version of FIII fragment, FIIIR.

All hamster-adapted re-encoded strains (rTs1h, rTs2h, rTs3h, rUAh, rCGh, rSSh, rNh and rTCGh) were obtained by replacing the WT FI fragment with the *Ap7M* FI fragment, FI*Ap7*.

Production and recovery. For producing each of the different YFV strains described in this study, 3 overlapping DNA fragments were synthesized *de novo* (Genscript) and amplified by High Fidelity PCR using the Platinum PCR SuperMix High Fidelity kit (Life Technologies) and specific sets of primers. After amplification, all DNA fragments were purified using Monarch PCR & DNA Cleanup kit 5µg (BioLabs) according to the manufacturer's instructions. Details regarding the subgenomic DNA fragments and the sets of primers used for the amplification step are available in the supplementary Table S6. The different subgenomic fragments combinations are described in Fig 1.

A final amount of 1µg of DNA (equimolar mix of subgenomic cDNA fragments) was transfected using Lipofectamine 3000 (Life Technologies) in a 25 cm2 culture flask of subconfluent cells containing 1mL of culture medium without antibiotics. The cell supernatants were harvested at

9 days post-transfection (dpi), aliquoted and stored at -80 °C. Each virus was then passaged two times onto Vero (*Asibi*-derived) or BHK21 (*Ap7M*-derived) cells and one last time onto BHK21 cells. Clarified cell supernatants from the second passage (virus stocks) were stored and used to perform viral RNA quantification, TCID50 assays and whole-genome sequencing (see corresponding sections).

Nucleic acids extraction

Samples (liver homogenate or cell culture supernatant) were extracted using either EZ1 Biorobot (EZ1 Virus Mini kit v2) or the QiaCube HT device (CadorPathogen kit) both from Qiagen. Inactivation was performed using either 200µL of AVL buffer (EZ1) or 100µL of VXL buffer and 100µL HBSS (Qiacube) according to the manufacturer's instructions.

Quantitative real-time RT-PCR assays

All quantitative real-time PCR (qRT-PCR) assays were performed using the EXPRESS SuperScript kit for One-Step qRT-PCR (Invitrogen) and the GoTaq® Probe 1-Step RT-qPCR System (Promega). Primers and probe sequences are detailed in the supplementary Table S6.

The amount of viral RNA was calculated from standard curves using a synthetic RNA transcript (concentrations of 10^7, 10^6, 10^5 and 10^4 copies/µL). Results were normalised using amplification (qRT-PCR) of the housekeeping gene actin (as described by Piorkowski and colleagues (5)). The standard curves generated for all the YFV-specific assays had coefficient of determination values (R2) >0.98 and amplification efficiencies were between 93% and 100%.

Competition assays

WT virus was grown in competition with each one of the 9, *Asibi*-derived, re-encoded viruses using two RNA ratios in triplicate (WT/re-encoded virus: 10/90 and 50/50). A global estimated MOI of 0.5 was used for the first inoculation. Viruses from each experiment were then passaged nine times with an incubation time of 3 days. At each passage, the estimated MOI was bottlenecked at approximately 0.6. After each infection (3 dpi), nucleic acids were extracted from the clarified culture supernatant using the QiaCube HT device (see corresponding section) and used for RT-PCR amplification. The detection of each virus was achieved using next generation sequencing (NGS) methods (see corresponding section), by amplifying a 256 nucleotides region and counting virus-specific patterns within the read population.

In vivo experiments

The laboratory hamster model reproduces well the features of human YF but requires the use of adapted strains (7). We recently implemented a hamster model for YFV, based on the use of a strain equivalent to the *Asibi*/hamster p7 strain described by Mac Arthur and colleagues in 2003 (8), *Ap7M*. When inoculated to hamsters, this strain induces a lethal viscerotropic disease similar to that described for YF *Ap7* virus in terms of (i) clinical signs of illness, (ii) weight evolution, (iii) viral loads in the liver and (iv) lethality (100%). This strain was derived from *Asibi* by including 10 mutations into the FI subgenomic region used for viral recovery with the ISA method. As we chose to target the FII subgenomic region for reencoding, the *Ap7M* strain was very well suited for testing the effects of reencoding on the biological properties of viruses *in vivo*.

Three-week-old female Syrian Golden hamsters were inoculated intra-peritoneally with 5.10^5 TCID50 of virus in a final volume of 100µL HBSS. In all experiments, a control group of 2 hamsters was kept uninfected. The clinical course of the viral infection was monitored by

following (i) the clinical manifestation of the disease (brittle fur, dehydration, prostration and lethargy), (ii) weight evolution and (iii) death. Weight was expressed as a normalised percentage of initial weight (%IW) and calculated as follows:

$$\%IW = \frac{W_{dn}}{W_{d0}} - \%IW_m + 1$$

(W_{dn}: weight at day n; W_{d0}: weight at the day of the inoculation or challenge; IW_n: mean of the *%IW* for control hamsters).

Liver samples were obtained from euthanised hamsters, grounded and treated with proteinase K (PK) before nucleic acid extraction using either the EZ1 Biorobot or the QiaCube HT device (see corresponding section). Serum was recovered from euthanised hamsters and stored (-80°C).

Serum neutralization assay

Sera were incubated for 20min at 56°C prior to viral serology. For each serum, serial 5-fold dilutions (first dilution: 1:250) of serum were tested on BHK21 cells for *Ap7M* strain infection inhibition. The plates were incubated for 44 hours before nucleic acid extraction. Virus quantification was achieved using the YFV-specific qRT-PCR system as described above. Then, a percent inhibition was calculated for each well as follows:

$$1 - \frac{qYFV_w}{qYFV_m}$$

With $qYFV_w$ the number of RNA copies in the analyzed well and $qYFV_m$ the mean number of RNA copies in the negative control wells. The IC50 values were determined from plots of percent inhibition versus log inhibitor concentration and calculated by non-linear regression

analysis using Graphpad PRISM software (v7.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com).

Whole and partial genome sequencing

Nucleotide sequences were determined using NGS methods: overlapping amplicons spanning either the complete genome sequence or a 256 nucleotides region (nucleotide positions: 4391-4646) were produced from the extracted RNA using the SuperScript® III One-Step RT-PCR System with Platinum®Taq High Fidelity kit (Invitrogen) and specific primers (detailed in the supplementary Table S7). Sequencing was performed using the PGM Ion torrent technology (Thermo Fisher Scientific) following the manufacturer's instructions.

Sequence determination. CDS consensus sequence determination was done using CLC genomics workbench software (CLC bio-Qiagen). Substitutions with a frequency higher than 5% were considered for the analysis of intra-population genetic diversity and major/minor variants identification (minor variants: 5%< variants frequency <=75%; major variants: variants frequency >75%).

Variants quantification. After sequencing of the 256 nucleotides region using the PGM Ion torrent technology. Automated read datasets provided by Torrent software suite 5.0.2 were trimmed according to quality score using CLC genomics workbench software (CLC bio-Qiagen) and 6 re-encoded or wild-type specific patterns (see supplementary Tables S8, S9 and S10) were counted within the read datasets using an in-house software. As a control, 2 patterns common to both viruses were counted to evaluate the total amount of virus for each sample. Pattern counts are given in the supplementary Table S11.

Statistical analysis

Viral loads comparisons were achieved using Wilcoxon rank-sum test and Kaplan-Meier survival analysis, using Mandel-Cox's Logrank tests. Both analyses were performed using R software (6). P-values below 0.05 were considered as significant.

Aknowledgements

We thank Morgan Seston, Ludivine Molina and Karine Barthélémy from UMR EPV, "Émergence des Pathologies Virales" for their technical assistance.

Funding

This work was supported by the European Virus Archive goes Global, http://global.europeanvirus-archive.com/ (European Union's Horizon 2020 research and innovation programme under grant agreement no. 653316); and the Agence Nationale de la Recherche, http://www.agence-nationalerecherche.fr/ (grant ANR-14CE14-0001 RNA Vacci-Code). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

References

1. Simmonds P. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res Notes. 2012;5:50.
2. King A. AM aLE. Virus Taxonomy. Ninth report of the International Committee on Taxonomy of Viruses.; 2014.
3. Zanotto PM, Gao GF, Gritsun T, Marin MS, Jiang WR, Venugopal K, et al. An arbovirus cline across the northern hemisphere. Virology. 1995;210(1):152-9.
4. Grard G, Moureau G, Charrel RN, Lemasson JJ, Gonzalez JP, Gallian P, et al. Genetic characterization of tick-borne flaviviruses: new insights into evolution, pathogenetic determinants and taxonomy. Virology. 2007;361(1):80-92.
5. Gould EA, Solomon T. Pathogenic flaviviruses. Lancet. 2008;371(9611):500-9.
6. Gaunt MW, Sall AA, de Lamballerie X, Falconar AK, Dzhivanian TI, Gould EA. Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography. J Gen Virol. 2001;82(Pt 8):1867-76.
7. Moureau G, Cook S, Lemey P, Nougairede A, Forrester NL, Khasnatinov M, et al. New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. PLoS One. 2015;10(2):e0117849.
8. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol. 1998;72(1):73-83.
9. Ng WC, Soto-Acosta R, Bradrick SS, Garcia-Blanco MA, Ooi EE. The 5' and 3' Untranslated Regions of the Flaviviral Genome. Viruses. 2017;9(6).
10. Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL, Strauss JH. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science. 1985;229(4715):726-33.
11. Bell JR, Kinney RM, Trent DW, Lenches EM, Dalgarno L, Strauss JH. Amino-terminal amino acid sequences of structural proteins of three flaviviruses. Virology. 1985;143(1):224-9.
12. Monath TP, Barrett AD. Pathogenesis and pathophysiology of yellow fever. Adv Virus Res. 2003;60:343-95.
13. Monath TP. Treatment of yellow fever. Antiviral Res. 2008;78(1):116-24.
14. Germain M, Cornet M, Mouchet J, Herve JP, Robert V, Camicas JL, et al. [Sylvatic yellow fever in Africa recent advances and present approach (author's transl)]. Med Trop (Mars). 1981;41(1):31-43.
15. Carrington CV, Auguste AJ. Evolutionary and ecological factors underlying the tempo and distribution of yellow fever virus activity. Infect Genet Evol. 2013;13:198-210.
16. Ellis BR, Barrett AD. The enigma of yellow fever in East Africa. Rev Med Virol. 2008;18(5):331-46.
17. Hanley KA, Monath TP, Weaver SC, Rossi SL, Richman RL, Vasilakis N. Fever versus fever: the role of host and vector susceptibility and interspecific competition in shaping the current and future distributions of the sylvatic cycles of dengue virus and yellow fever virus. Infect Genet Evol. 2013;19:292-311.
18. Mir D, Delatorre E, Bonaldo M, Lourenco-de-Oliveira R, Vicente AC, Bello G. Phylodynamics of Yellow Fever Virus in the Americas: new insights into the origin of the 2017 Brazilian outbreak. Sci Rep. 2017;7(1):7385.
19. Ahmed QA, Memish ZA. Yellow fever from Angola and Congo: a storm gathers. Trop Doct. 2017;47(2):92-6.
20. Kraemer MU, Faria NR, Reiner RC, Jr., Golding N, Nikolay B, Stasse S, et al. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015-16: a modelling study. Lancet Infect Dis. 2017;17(3):330-8.
21. Theiler M, Smith HH. The Effect of Prolonged Cultivation in Vitro Upon the

Pathogenicity of Yellow Fever Virus. J Exp Med. 1937;65(6):767-86.
22. Barrett MGS. Yellow fever vaccine. In: Elsevier, editor. Vaccines: Sixth Edition2012. p. 870-968.
23. Monath TP. Review of the risks and benefits of yellow fever vaccination including some new analyses. Expert Rev Vaccines. 2012;11(4):427-48.
24. Ferguson M, Shin J, Knezevic I, Minor P, Barrett A, Group WHOW. WHO Working Group on Technical Specifications for Manufacture and Evaluation of Yellow Fever Vaccines, Geneva, Switzerland, 13-14 May 2009. Vaccine. 2010;28(52):8236-45.
25. Monath TP, Vasconcelos PF. Yellow fever. J Clin Virol. 2015;64:160-73.
26. Monath TP. Yellow fever vaccine. Expert Rev Vaccines. 2005;4(4):553-74.
27. Sabin AB, Boulger LR. History of Sabin attenuated poliovirus oral live vaccine strains. Journal of Biological Standardization. 1973;1(2):115-8.
28. Schwarz AJ. Preliminary tests of a highly attenuated measles vaccine. Am J Dis Child. 1962;103:386-9.
29. Ikic D, Juzbasic M, Beck M, Hrabar A, Cimbur-Schreiber T. Attenuation and characterisation of Edmonston-Zagreb measles virus. Ann Immunol Hung. 1972;16(0):175-81.
30. Hilleman MR, Buynak EB, Weibel RE, Stokes J, Jr., Whitman JE, Jr., Leagus MB. Development and evaluation of the Moraten measles virus vaccine. JAMA. 1968;206(3):587-90.
31. Buynak EB, Hilleman MR. Live attenuated mumps virus vaccine. 1. Vaccine development. Proc Soc Exp Biol Med. 1966;123(3):768-75.
32. Minor PD. The polio-eradication programme and issues of the end game. J Gen Virol. 2012;93(Pt 3):457-74.
33. Wang E, Ryman KD, Jennings AD, Wood DJ, Taffs F, Minor PD, et al. Comparison of the genomes of the wild-type French viscerotropic strain of yellow fever virus with its vaccine derivative French neurotropic vaccine. J Gen Virol. 1995;76 (Pt 11):2749-55.
34. Lukashev AN, Lashkevich VA, Ivanova OE, Koroleva GA, Hinkkanen AE, Ilonen J. Recombination in circulating enteroviruses. J Virol. 2003;77(19):10423-31.
35. Liu J, Zhang H, Zhao Y, Xia L, Guo C, Yang H, et al. Characterization of four vaccine-related polioviruses including two intertypic type 3/type 2 recombinants associated with aseptic encephalitis. Virol J. 2016;13(1):162.
36. Burns CC, Shaw J, Campagnoli R, Jorba J, Vincent A, Quay J, et al. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J Virol. 2006;80(7):3259-72.
37. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. J Virol. 2006;80(19):9687-96.
38. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 1985;2(1):13-34.
39. Shabalina SA, Ogurtsov AY, Spiridonov NA. A periodic pattern of mRNA secondary structure created by the genetic code. Nucleic Acids Res. 2006;34(8):2428-37.
40. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science. 2009;324(5924):255-8.
41. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat Genet. 2011;43(3):242-5.
42. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 2012;22(6):1059-68.
43. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by

932 codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies.
933 Elife. 2014;3:e04531.
934 44. Jimenez-Baranda S, Greenbaum B, Manches O, Handler J, Rabadan R, Levine A, et al.
935 Oligonucleotide motifs that disappear during the evolution of influenza virus in humans
936 increase alpha interferon secretion by plasmacytoid dendritic cells. J Virol. 2011;85(8):3893-
937 904.
938 45. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene
939 mimicry in influenza and other RNA viruses. PLoS Pathog. 2008;4(6):e1000079.
940 46. Nougairede A, De Fabritus L, Aubry F, Gould EA, Holmes EC, de Lamballerie X.
941 Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus
942 in primate and mosquito cells. PLoS Pathog. 2013;9(2):e1003172.
943 47. de Fabritus L, Nougairede A, Aubry F, Gould EA, de Lamballerie X. Attenuation of tick-
944 borne encephalitis virus using large-scale random codon re-encoding. PLoS Pathog.
945 2015;11(3):e1004738.
946 48. Martinez MA, Jordan-Paiz A, Franco S, Nevot M. Synonymous Virus Genome
947 Recoding as a Tool to Impact Viral Fitness. Trends Microbiol. 2016;24(2):134-47.
948 49. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. Codon usage between genomes
949 is constrained by genome-wide mutational processes. Proc Natl Acad Sci U S A.
950 2004;101(10):3480-5.
951 50. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, et al.
952 Architecture and secondary structure of an entire HIV-1 RNA genome. Nature.
953 2009;460(7256):711-6.
954 51. Thurner C, Witwer C, Hofacker IL, Stadler PF. Conserved RNA secondary structures in
955 Flaviviridae genomes. J Gen Virol. 2004;85(Pt 5):1113-24.
956 52. Aubry F, Nougairede A, de Fabritus L, Querat G, Gould EA, de Lamballerie X. Single-
957 stranded positive-sense RNA viruses generated in days using infectious subgenomic amplicons.
958 J Gen Virol. 2014;95(Pt 11):2462-7.
959 53. Greenbaum BD, Rabadan R, Levine AJ. Patterns of oligonucleotide sequences in viral
960 and host cell RNA identify mediators of the host innate immune system. PLoS One.
961 2009;4(6):e5969.
962 54. Karlin S, Doerfler W, Cardon LR. Why is CpG suppressed in the genomes of virtually
963 all small eukaryotic viruses but not in those of large eukaryotic viruses? J Virol.
964 1994;68(5):2889-97.
965 55. Rima BK, McFerran NV. Dinucleotide and stop codon frequencies in single-stranded
966 RNA viruses. J Gen Virol. 1997;78 (Pt 11):2859-70.
967 56. Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids
968 Res. 1980;8(7):1499-504.
969 57. Kow YW. Repair of deaminated bases in DNA. Free Radic Biol Med. 2002;33(7):886-
970 93.
971 58. Simmen MW. Genome-scale relationships between cytosine methylation and
972 dinucleotide abundances in animals. Genomics. 2008;92(1):33-40.
973 59. Adams AP, Travassos da Rosa AP, Nunes MR, Xiao SY, Tesh RB. Pathogenesis of
974 Modoc virus (Flaviviridae; Flavivirus) in persistently infected hamsters. Am J Trop Med Hyg.
975 2013;88(3):455-60.
976 60. Johnson HN. Long-term persistence of Modoc virus in hamster-kidney cells. In vivo
977 and in vitro demonstration. Am J Trop Med Hyg. 1970;19(3):537-9.
978 61. Leyssen P, Croes R, Rau P, Heiland S, Verbeken E, Sciot R, et al. Acute encephalitis, a
979 poliomyelitis-like syndrome and neurological sequelae in a hamster model for flavivirus
980 infections. Brain Pathol. 2003;13(3):279-90.
981 62. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ, Nicol MQ, et al. Elevation of CpG

frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. Elife. 2016;5:e12735.
63. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. Nucleic Acids Res. 2014;42(7):4527-45.
64. Witteveldt J, Martin-Gans M, Simmonds P. Enhancement of the Replication of Hepatitis C Virus Replicons of Genotypes 1 to 4 by Manipulation of CpG and UpA Dinucleotide Frequencies and Use of Cell Lines Expressing SECL14L2 for Antiviral Resistance Testing. Antimicrob Agents Chemother. 2016;60(5):2981-92.
65. ter Meulen J, Sakho M, Koulemou K, Magassouba N, Bah A, Preiser W, et al. Activation of the cytokine network and unfavorable outcome in patients with yellow fever. J Infect Dis. 2004;190(10):1821-7.
66. Quaresma JA, Pagliari C, Medeiros DB, Duarte MI, Vasconcelos PF. Immunity and immune response, pathology and pathologic changes: progress and challenges in the immunopathology of yellow fever. Rev Med Virol. 2013;23(5):305-18.
67. McArthur MA, Suderman MT, Mutebi JP, Xiao SY, Barrett AD. Molecular characterization of a hamster viscerotropic strain of yellow fever virus. J Virol. 2003;77(2):1462-8.
68. Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat Struct Mol Biol. 2013;20(2):237-43.
69. Simmonds P, Tulloch F, Evans DJ, Ryan MD. Attenuation of dengue (and other RNA viruses) with codon pair recoding can be explained by increased CpG/UpA dinucleotide frequencies. Proc Natl Acad Sci U S A. 2015;112(28):E3633-4.
70. Shen SH, Stauft CB, Gorbatsevych O, Song Y, Ward CB, Yurovsky A, et al. Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. Proc Natl Acad Sci U S A. 2015;112(15):4749-54.
71. Martrus G, Nevot M, Andres C, Clotet B, Martinez MA. Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. Retrovirology. 2013;10:78.

Supporting information

S1 Supplementary Protocols.

S2 Supplementary Results.

S1 Table. Yellow Fever dataset: sequence accession numbers.

S2 Table. Rare codons identified within a complete alignment of 35 YF sequences.

S3 Table. Putative secondary strucures identified within YFV subgenomic fragment FII (position 3924 to 6759).

S4 Table. Ecological groups and sequence accession numbers used for vector/host preference associated patterns identification.

S5 Table. Trinucleotides frequencies amongst ISF, MBF and NKV groups. Each frequency was expressed as the ratio between the trinucleotide count and the total number of trinucleotides in the considered sequence. The differences in frequencies were then calculated and summed between ISF and both MBF and NKV groups and between MBF and both NKV and ISF groups.

S6 Table. PCR/qRT-PCR primers and Yellow fever subgenomic fragments description.

S7 Table. RT-PCR primers used for pre-sequencing amplification.

S8 Table. Universal, wild-type and reencoded-specific patterns (1).

S9 Table. Universal, wild-type and reencoded-specific patterns (2).

S10 Table. Universal, wild-type and reencoded-specific patterns (3).

S11 Table. Wild-type and reencoded-specific patterns counts at 10th competition passage. For each WT/re-encoded virus ratio and each re-encoded virus pattern counts were calculated as a mean pattern count for the 3 replicates. For each replicate the pattern count is the mean count for the three virus-specific patterns. Each specific pattern count is the sum of the sense and reverse patterns counts within a given read population (obtained through NGS sequencing).

S12 Table. Mature proteins positions within Yellow Fever Virus polyprotein sequence. (obtained from UniProt database, www.uniprot.org)