Exploratory re-encoding of Yellow Fever Virus genome: analysis of *in vitro* and *in vivo* replicative phenotypes

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 an innovative approach for generating live-attenuated vaccine candidates under a defined set of rules that can be applied to any RNA virus. The 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. We revisited this concept and set up an exploratory work to identify the mutations that generate the least deleterious impact on the replicative fitness of yellow fever virus (YFV). Using comparative bioinformatic analyses of genomes from the *Flavivirus* genus, we defined several types of synonymous mutations according to their hypothetical "fitness cost" based on frequency, genome location and impact on secondary RNA structure. The re-encoded sequences were generated with 100 to 400 mutations in the NS2A to NS4B coding region of YF *Asibi* and *Ap7M* genomes. After being rescued using the Infectious Subgenomic Amplicons (ISA) method, parent and re-encoded viruses were compared in mammalian cells (competition experiments) and in a hamster model (infection and immunisation experiments), combined to complete genome sequencing and intra-population diversity analysis. None of the re-encoded viruses but the highly re-encoded FII/IIIR strain (741 re-encoded sites) showed significant decrease in replicative fitness *in vitro*. Whilst some re-encoded strains showed a phenotype close to the hamster-virulent strain *Ap7M*, most of them exhibited attenuated phenotypes *in vivo*, with alleviated virulence and sometimes, decreased replicative fitness. This allowed us to identify new types of mutations endowed with a very limited impact on YFV replicative fitness. In addition, we observed that some of the re-encoded strains could induce a robust, protective immunity in hamsters upon challenge with *Ap7M* virus. Altogether, these results may allow to improve procedures for viral attenuation through synonymous codon replacement.

Author Summary

Genome re-encoding is an innovative approach for live-attenuated virus production that has already proved to be effective against several major human pathogens. It is based on a set of rules that can be applied to any RNA virus, but the mechanisms at stake in the attenuation process remain to be fully characterized. The yellow fever virus (YFV; genus Flavivirus), is an arthropod-borne human pathogen that causes severe viscerotropic disease in association with high fever in human populations from the tropical regions of Africa and South-America. We conducted an exploratory work using the YF *Asibi* strain to identify the re-encoding strategies that may be the best suited for live-attenuated virus generation. Based on bioinformatic analyses of genomes from the *Flavivirus* genus, we defined several re-encoding strategies and designed thirteen YF re-encoded strains. We rescued both parent (*Asibi*, *Ap7M*) and re-encoded strains using the pioneering Infectious Subgenomic Amplicons (ISA) reverse-genetics method before characterizing and comparing the viruses *in vitro* and *in vivo*. Beyond the efficient attenuation of YFV *in vivo*, our results provide new insights into the mechanisms involved into the attenuation process and may be most useful for the future development of efficient, safe and cost-effective live-attenuated vaccines with re-encoded genomes.

Introduction

The *Flavivirus* genus (family *Flaviviridae)* brings together 53 taxonomically recognized species and at least 19 “tentative” species (1). This ensemble is endowed with an exceptional diversity in ecological networks that widely correlates with the phylogenetic relationships within the genus (2-5). To date, a majority of identified flaviviruses are arthropod-borne viruses (arboviruses), whose 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”) (6, 7). 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 are 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 (8). ORF translation gives rise to a polyprotein that is clived co- and post-translationally into 3 structural (C-prM-E) and 7 non-structural proteins (NS1-2A-2B-3-4A-4B-5) (1).

YFV is the funding member of the genus *Flavivirus*, that owns 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 (9). 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 (10). YFV natural upkeep is conditioned by the presence of both its non-human primate (NHP) host(s) and mosquito vector(s), between which viral transmission occurs in a sylvatic cycle. Sylvatic maintenance holds in the tropical regions of

both the African and South-American continents, where human cases result in the first place from NHP to human transmission through mosquito bites (11-15), as recently illustrated in Brazil during the 2016-2017 outbreak (15). In some instances, in Africa, large interhuman transmission cycles occur, leading to epidemics that can reach wide extents, of which the latest was described in Angola and Democratic Republic of Congo between 2015 and 2016 (16, 17).

YFV live-attenuated vaccine (strain 17D) was developed in 1936 by M.Theiler (18). It served as early as 1938 in Brazil (19, 20) and two of its substrains (17DD and 17D-204) are still widely used for vaccine manufacturing (21). Throughout decades of use of 17D vaccine, with 20 to 60 million doses distributed annually (22), YF vaccination has proven to be safe and efficient, providing a long-lasting immunity (23) with rare adverse events (22). The YF 17D strain was obtained through serial passages (>200) of the wild-type (WT) *Asibi* strain in mouse and chicken embryo tissues (23). Similar empirical methods allowed the production of several notable live-attenuated vaccines including those against Poliovirus (24), Measles (25-27) and Mumps (28). This strategy relies on attenuation mechanisms that involve a limited number of non-synonymous mutations. The latter are subject to (i) reversion, as described in the case of the poliovirus vaccine (29); (ii) generation of new biological properties, as illustrated by the gain of neurovirulence observed for the YFV French neurotropic vaccine strain (30); and (iii) recombination between vaccine strains, as documented for the poliovirus vaccine, that includes three different strains (29).

A new, promising, approach for virus attenuation called codon re-encoding, has been developed a decade ago by Burns (31) and Mueller (32). 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 and/or recombination between vaccine strains, whilst the use of silent mutations significantly lowers the risk 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 (33). Several other mechanisms have been proposed since then, including implication in mRNA structure and folding (34, 35), microRNA-targeting (36, 37) and enhanced recognition by the immune system (38), that may vary according to the host (39, 40). 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 (31, 32, 41-43). Considering a major role of genome-wide mutational processes in the shaping of synonymous sites (44), 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 on synonymous codon use (41, 42, 45, 46).

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 categorised silent mutations and sought to compare the respective impact of the different groups on the replicative phenotype and virulence of YFV. Our final aim was to identify the mutation types endowed with the least deleterious effect on viral replicative fitness, which could in the future 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 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.*, manuscript submitted) derived from the previously described hamster-virulent YF *Asibi*/hamster p7 strain (47). When inoculated to hamsters, both strains strain 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 along with a hypothetical “fitness cost” and used them for the design of 13 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, complete genome sequencing and analysis of intra-population genetic diversity. Although no significant attenuation was observed *in vitro*, we reported various degrees of attenuation *in vivo*. In addition, we showed that the strains endowed with the most attenuated *in vivo* phenotypes could induce robust protective immunity in hamsters subsequently challenged with the *Ap7M* virus.

Results

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

***In silico* analysis.** Previous analyses on the impact of random codon re-encoding on chikungunya (CHIKV) and tick-borne encephalitis (TBEV) viruses, provided us valuable information regarding the extent of re-encoding needed to obtain observable *in vitro* and/or *in vivo* attenuation. Nougairède and colleagues observed a decrease of CHIKV replicative fitness *in vitro*, using between 264 and 882 mutations located within the nsp1, nsp4 and/or envelope proteins coding regions (41). 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 (42). Based on those results, we used 300 to 350 mutations for the design of our re-encoded viruses and defined a framework for the design of re-encoded strains. First, we chose a target mutagenesis region located within the coding sequences of proteins NS2A to NS4B. Therefore, neither the structural proteins nor the viral RNA-dependent RNA polymerase coding sequences were affected by re-encoding. Besides, this design had the advantage of conveniently targeting an ISA fragment without additional non-viral sequences. To reduce the potential deleterious side-effects of re-encoding, 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 “mutable sites”. We also ~~also~~ avoided creating or removing any rare codon during the re-encoding process (rare codons are defined in the supplementary protocol).

In an *in silico* analysis, using either YFV or flaviviral genomes datasets, we defined several types of synonymous mutations (see below and in the supplementary results, where results of the bioinformatic analysis are detailed). We defined “ubiquitous” mutations as those precisely defined on an objective molecular basis: this category includes transitions, transversions, and

transitions on potential secondary structures. By contrast, we included in "specific" mutations the 4 mutations types associated to the modification of vector/host preferential patterns identified within the *Flavivirus* genus. This included introduction of dinucleotides (CpG and UpA) and trinucleotides (TCG and ACA). CpG and UpA dinucleotides are commonly low in vertebrate cells, TCG trinucleotides are found in higher proportion in insect-specific flaviviruses, whilst ACA trinucleotides are more frequent in mosquito-borne flaviviruses. Based on observed frequencies, either a “high” or a “low” potential fitness cost was associated to each mutation type, according to the hypothesis that the least deleterious a mutation was, the most frequently it should be observed within the YFV CDS alignment (a detailed classification based on mutation type and fitness cost is available in Table 1).

	Low deleterious impact	High deleterious impact
Ubiquitous	-Simple transitions -Transitions on sites potentially involved in secondary structures	-Simple transversions
Specific	-Transitions leading to ACA trinucleotides introduction	-Transitions leading to CpG/UpA dinucleotides introduction -Transitions leading TCG trinucleotides introduction

Table 1. *Mutations categories.* Based on bioinformatical analysis, 6 types of mutations were classified according to their hypothetical fitness cost.

In total, 14 re-encoded viruses were designed based on the *Asibi* strain sequence. The re-encoded viruses n1 to n11, n14 and n15 were designed with 100 to 400 mutations located between positions 3,924 and 6,759 of YF *Asibi* complete coding sequence (CDS). The re-encoded strain FII/IIIR included 388 additional mutations located between positions 6,846 and 9,765. Finally, 13 hamster-adapted re-encoded strains were created from viruses n1 to n11, n14 and n15 by introducing exactly the same mutations in the CDS of the hamster-virulent YFV *Ap7M* virus. Together with *Asibi* and *Ap7M* strains, those 27 re-encoded strains were produced to allow the comparison, *in cellulo* and *in vivo*, of the phenotypes of the wild-type (WT) and re-encoded viruses.

Design of re-encoded strains including only “ubiquitous” mutations. Re-encoded strains YFV n1, n2 and n3, were designed by introducing increasing numbers of “simple”

transitions (Ts) in the CDS of the reference strain *Asibi* (AY640589). This refers to synonymous mutations located outside potential secondary structures sites that do not create new CpG or UpA dinucleotides. Strain YFV FII/IIIR included all mutations of YFV n3 and 388 additional transitions located between position 6,846 and 9,765 of the CDS, with no introduction of CpG/UpA dinucleotides (741 transition mutations in total).

Strain YFV n10 was re-encoded using 339 Ts of which 50 were located at sites corresponding to potential secondary structures into the reference sequence.

Strains YFV n4 and n5 combined simple transitions (Ts) and simple transversions (Tv) with Tv/Ts ratios of 17% and 52% for n4 and n5 strains, respectively.

Design of re-encoded strains including only "specific" mutations. Viruses YFV n6 to n9 were used to investigate the effect of UpA and CpG dinucleotides introduction on the viral replicative phenotype. Transitions leading to UpA dinucleotide introduction (UpA-Ts) were combined to Ts in varying proportions for strains n6 (5% of "UpA-Ts) and n7 (18% of UpA-Ts). Similarly, strains n8 and n9 involved different proportions of CpG-Ts (7% and 18%, respectively).

For appraising the impact of introducing vector/host preferential patterns, transitions creating TCG and/or ACA trinucleotides (TCG-Ts and ACA-Ts, respectively) were used for the design of re-encoded strains YFV n14 (118 TCG-Ts) and n15 (118 TCG-Ts and 45 ACA-Ts). Because of the scarcity of mutable sites allowing TCG or ACA introduction, design rules were relaxed: both transitions and transversions were used and some mutations affected sites corresponding to potential secondary structures (12 and 20 for strains n14 and n15, respectively) as well as sites for which no mutation had been observed within the 35 sequences alignment (11 and 14 for strains n14 and n15, respectively).

Name	Description	Transitions	Transversions	CpG-Ts	UpA-Ts	SII-Ts	TCG-Ts	ACA-Ts	Transitions on NM	Total	Titre (TCID50/mL)
nA	*Asibi*	0	0	0	0	0	0	0	0	0	1E+08
n1	Transitions-low	101	0	0	0	0	2	2	0	101	4E+05
n2	Transitions-intermediate	247	0	0	0	0	2	7	0	247	4E+07
n3	Transitions-high	353	0	0	0	0	3	10	0	353	4E+06
n4	Transversions- (Tv/Ts=17)	303	50	0	0	0	2	15	0	353	3E+08
n5	Transversions-(Tv/Ts=52)	232	120	0	0	0	1	21	0	352	2E+07
n6	Transitions to UpA-5%	336	0	0	18	0	3	6	0	336	3E+06
n7	Transitions to UpA-18%	326	0	0	59	0	2	10	0	326	3E+06
n8	Transitions to CpG-7%	344	0	23	0	0	11	10	0	344	3E+06
n9	Transitions to CpG-18%	325	0	59	0	0	19	9	0	325	3E+06
n10	Transitions on potential secondary structures	339	0	0	0	50	3	12	0	339	7E+06
n11	Combinatory strategy	179	54	59	83	0	17	3	0	233	4E+06
n14	Transitions to TCG	46	72	111	0	12	118	0	11	118	2E+06
n15	Transitions to TCG and ACA	60	103	111	8	20	118	45	14	163	6E+06
FII/III Reenc	Transitions on 2nd and 3rd fragments	741	0	0	0	0	7	15	0	741	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 potential secondary structures; TCG-Ts: transition leading to TCG trinucleotide introduction; ACA-Ts: transition leading to ACA trinucleotide introduction; NM: Site for which no mutation was reported in the 35 YFV CDS alignment. Viral titres are given in TCID50/mL.

Finally, a combinatory strategy was used for YFV strain n11 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.

All details regarding re-encoded strains design are available in Table 2. Importantly, all the mutations included in re-encoded variant strains were synonymous and were exclusively introduced at mutable sites (see definition in the *in silico* analysis).

Production of YFV re-encoded strains.

All strains used in this study (1 WT, 1 hamster-adapted and 27 re-encoded) were produced using the ISA method (3) that allows recovering infectious viruses after transfection into permissive cells of overlapping subgenomic DNA fragments covering the entire genome. All fragments combination schemes used for virus production are detailed in Fig 1.

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 (3). The hamster-virulent strain YFV *Ap7M*, was derived 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.*, submitted manuscript).

Re-encoded strains 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 YFV

FII/IIIReenc strain, the WT fragment FI was combined to the re-encoded FII fragment of strain n3 and to a re-encoded version of FIII fragment, FIIIR.

All hamster-adapted re-encoded strains (n1h to n11h, n14h and n15h) were obtained by replacing the WT FI fragment with the *Ap7M* FI fragment, FI*Ap7*.

All strains were recovered after amplification and transfection of the 3 corresponding subgenomic amplicons in BHK21 cells according to the ISA method (3). The viral culture supernatant was passaged twice onto Vero or BHK21 cells and a third time onto BHK21 cells before storage, next-generation sequencing and either *in vitro* or *in vivo* study.

In vitro behaviour of YFV re-encoded strains

Sequence analysis. Next Generation Sequencing (NGS) was performed on the cell culture supernatants (viral stocks). Both the integrity of the genome and intra-population genetic diversity could be assessed on the CDS of all viruses. For intra-population genetic diversity analysis, a variant was regarded as major when the corresponding nucleotide proportion (CNP) was over 75%. 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 a result of the adaptation of re-encoded strains to culture conditions. Overall, low numbers of mutations associated to major variants were observed, they ranged from 0 to 6, with an average number of 2 changes/virus and an average 44% of NS mutations. Mutations were equally distributed between re-encoded and non-re-encoded regions. Apart from one mutation (see below), no convergence was observed between viral sequences, suggesting a limited adaptation of mutant strains to culture conditions, in coherence with the low number of *in cellulo* passages achieved (3). The total

Strain name	Reencoded region	Position (nt)	Protein	Mutation	Expected sequence (nt)	Observed sequence (nt)	Observed sequence (2nd variant (nt))	1st variant proportion	2nd variant proportion
YFV Asibi	Native strain	2455	NS1	K41E	A	G	A	67,6%	32,4%
		5796	-	-	A	G	A	67,8%	32,2%
		7040	NS4B	I91T	T	C	T	67,6%	32,4%
YFV n1	3927-6753	1899	-	-	T	C	T	51,8%	48,2%
		1999	Env	D382N	G	A	G	62,6%	37,4%
		7465	NS4B	Y233H	T	C	T	52,0%	47,9%
		7509	-	-	T	C	T	51,2%	48,0%
YFV n2	3927-6753	2693	NS1	K120T	A	C	A	70,3%	29,5%
		4833	-	-	G	A	G	74,4%	25,6%
		5430	-	-	C	T	C	77,5%	22,3%
		5442	-	-	C	T	C	82,0%	18,0%
YFV n3	3927-6753	4357	NS2B	M99V	A	G	A	78,3%	21,7%
		9150	-	-	A	G	A	78,5%	21,5%
YFV n4	3927-6753	2630	NS1	R99K	G	A	G	63,7%	36,3%
		6882	-	-	T	A	T	63,4%	36,5%
		7443	-	-	A	G	A	63,5%	36,5%
		7577	NS4B	D270G	A	G	A	64,0%	35,9%
YFV n5	3927-6753	7053	NS4B	I95M	A	G	-	>95%	-
		7865	NS5	V116A	T	C	-	>95%	-
YFV n6	3927-6753	3976	NS2A	A196T	G	A	-	>95%	-
		7107	NS4B	I113M	A	G	A	53,1%	46,9%
YFV n7	3927-6753	-	-	-	-	-	-	-	-
YFV n8	3927-6753	675	-	-	G	A	-	>95%	-
		10153	-	-	T	C	-	>95%	-
YFV n9	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%	-
YFV n10	3939-6759	561	-	-	T	C	-	>95%	-
		3190	NS1	I286V	A	G	-	>95%	-
		4308	-	-	G	T	-	>95%	-
		4623	-	-	A	G	-	>95%	-
		6012	-	-	A	G	-	>95%	-
YFV n11	3924-6750	7018	NS4B	I84V	A	G	A	69,1%	30,8%
		8274	-	-	C	T	C	69,5%	30,5%
YFV n14	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%	-
YFV n15	3927-6759	231	-	-	C	T	C	70,3%	29,7%
		2606	NS1	Q91R	A	G	A	74,8%	25,1%
		6834	-	-	T	C	T	72,0%	28,0%
YFV FII IIIR	3927-9765	6846	NS4B	-	C	T	C	64,90%	35,1%

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.

VIRUS	Mutations	Non-Synonymous mutations	Variants	Reversions	Mutations in reencoded region	Variants in reencoded region
Asibi	0	0	12	0	0	1
n1	0	0	12	0	0	0
n2	2	0	9	2	2	3
n3	2	1	6	0	1	2
n4	0	0	9	0	0	3
n5	2	2	4	0	0	2
n6	1	1	5	0	1	0
n7	0	0	16	0	0	5
n8	2	0	4	0	0	1
n9	4	2	5	2	4	1
n10	5	1	1	0	3	0
n11	0	0	11	0	0	2
n14	6	2	4	0	0	1
n15	0	0	17	0	0	7
FII/IIIR	0	0	15	0	0	9
n1h	0	0	13	0	0	4
n2h	0	0	22	0	0	6
n3h	3	3	13	0	1	3
n4h	4	1	8	0	1	3
n5h	6	3	11	0	1	0
n6h	0	0	5	0	0	0
n7h	0	0	17	0	0	5
n8h	0	0	9	0	0	1
n9h	0	0	10	0	0	0
n10h	0	0	19	0	0	2
n11h	5	1	4	0	0	0
n14h	2	2	7	0	0	0
n15h	4	2	5	0	1	0

Table 4. *Sequence analysis of Asibi, 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*.

numbers of variants were heterogeneous, ranging from 1 to 17, with a mean of 8 variants per virus. Interestingly, variants were most frequently observed outside re-encoded regions. Only two viruses (YFV n2 and n9) showed mutation reversions, of which one was shared between the two viruses and corresponded to a simple transition (C to T), at CDS position 5,430. The others corresponded to simple transitions (C to T and A to G, respectively), at CDS positions 5,442 and 5,667 of YFV n2 and n9, respectively.

For strains YFV n3, n5, n6, n9, n10 and n14, non-synonymous mutations corresponding to major variants were observed, always in the non-structural proteins sequences.

Strain YFV n3 included a non-synonymous (NS) change within the NS2B protein (M99V). However, YFV FII/IIIR virus, derived from YFV n3 (*e.g.* produced using n3 FII-re-encoded fragments), did not include any NS mutation in its sequence. Therefore, it is unlikely that n3 NS change is due to the adaptation of the re-encoded virus to culture conditions.

The *Asibi* (wild-type) strain sequence did not arbour any NS change associated to major variant and nearly no convergence was observed between viral sequences. In addition, the viruses only went through a limited number of *in cellulo* passages. Hence, it is probable that the NS mutations observed in viruses YFV n5, n6, n9, n10 and n14 did not result from a selective pressure due to culture conditions but may be artefacts of the production method.

***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 (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 FII re-encoded strains, which infectious titres ranged between 2.10^6 and 3.10^8 $TCID_{50}$/mL. However, viruses YFV n1 was unexpectedly endowed with the lowest (4.10^5) infectious titre. When compared to *Asibi*, YFV FII/IIIR virus had a remarkably low infectious titre of 6.10^3 $TCID_{50}$/mL.

For each of the re-encoded virus, 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 FII re-encoded viruses, 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 countings are detailed in the supplementary Table S11). In contrast, YFV FII/IIIR virus showed a clear reduction in replicative fitness and could not be detected after the 5th passage.

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). No significant difference was observed between strain *Ap7M* (9.10^8 TCID50/mL) and most re-encoded viruses, which infectious titres ranged from 2.10^8 to 9.10^8 TCID50/mL. However, viruses YFV n11h and n14h, showed 1 to 2-log reduction in their viral titres (6.10^6 and 6.10^7 TCID50/mL, respectively).

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

As observed for *Asibi*-re-encoded viruses, the numbers of mutations associated to major variants were rare, ranging from 0 to 6, with an average number of 2 changes/virus and an 58% average of non-synonymous mutations. Mutations were mainly distributed outside re-encoded regions (83%). Furthermore, no reversion mutations were observed. The total number of variants per virus varied among strains, ranging from 4 to 22,

Mutant name	Titre (TCID50/mL)	Mortality	Mean viral loads (RNA copies/g of liver)			Mean normalized 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%
n1h	*6E+08*	0%	7E+08^{EP}	**5E+07^{A}**	3E+07	106%	92%
n2h	*6E+08*	17%	1E+10	6E+08	3E+09	107%	78%
n3h	*6E+08*	20%	9E+08	9E+08	2E+09	104%	78%
n4h	*2E+08*	0%	3E+10^{EP}	8E+09^{EP}	8E+07	105%	74%
n5h	*9E+08*	50%	6E+09	**3E+09^{A}**	3E+09	94%	55%
n6h	*4E+08*	60%	3E+10^{EP}	**8E+09^{A}**	4E+09	96%	65%
n7h	*4E+08*	0%	3E+09^{EP}	2E+09	**3E+07^{A}**	96%	78%
n8h	*6E+08*	40%	2E+10	1E+10^{EP}	2E+08	90%	69%
n9h	*4E+08*	0%	7E+08^{EP}	2E+09^{EP}	**5E+06^{A}**	91%	75%
n10h	*4E+08*	0%	1E+10	2E+09^{EP}	9E+07	89%	62%
n11h	*6E+06*	0%	1E+09^{EP}	**8E+07^{A}**	**1E+07^{A}**	97%	96%
n14h	*6E+07*	0%	4E+09	5E+08^{EP}	6E+07	97%	80%
n15h	*2E+08*	0%	5E+10^{EP}	1E+09^{EP}	8E+07	97%	67%

407

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 normalized percentage of initial weight means are given for all viruses. Viral loads, normalized 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) was 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) was indicated using both bold characters and an “A” superscript next to the corresponding viral loads’ mean.

with a mean number of 11 variants per virus. Only a minority of variants fell within the re-encoded regions (13%).

As observed for *Asibi*, the wild-type *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, it is likely that the NS mutations observed in viruses YFV n1h, n3h, n4h, n5h, n11h, n14h and n15h may rather be artefacts of the production method than the result of the adaptation of re-encoded viruses to culture conditions.

***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 normalized percentage of the initial weight (%IW)) and (iii) death. Respectively 3 and 4 hamsters were euthanized 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

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
N1H	489	-	-	A	G	-	g	a	1-4806, 5430-7566	A	-
	836	prM	A158V	C	T	A → V	C	-		T	-
	1601	Env	N249S	A	G	N → S	g	a		A	-
	1608	-	-	A	G	-	g	a		A	-
	2238	NS1	L79F	C	T	L → F	C	-		t	c
	2984	NS1	E217G	A	G	E → G	g	a		A	-
	3253	NS1	I307V	A	G	I → V	A	-		g	a
	4089	-	-	A	G	-	g	a		A	-
	4316	NS2B	K85R	A	G	K → R	g	a		A	-
	4574	NS3	G41A	G	C	G → A	G	-		C	g
	4746	-	-	A	G	-	g	a		A	-
	5442	-	-	C	T	-	C	-		T	-
	6273	-	-	T	C	-	c	t		T	-
	7107	NS4B	I113M	A	G	I → M	A	-		g	a
	9039	-	-	A	G	-	g	a		n/a	
N2H	-	-	-	-	-	-	-	-	1-3858, 4653-5771, 6083-10236	-	-
N3H	1649	Env	T265I	C	T	T → I	C	t	1-10236	T	-
	4604	NS3	R51M	T	G	R → M	G	t		T	g
	7668	2k	D50E	C	A	D → E	C	a		a	c
N4H	441	-	-	C	T	-	T	c	1-7674, 8782-9687	T	-
	655	prM	N98D	A	G	N → D	G	a		G	-
	2541	-	-	A	G	-	G	a		A	-
	3128	NS1	N265S	A	G	N → S	A	-		G	-
	3203	NS1	N290S	A	G	N → S	A	-		G	-
	5991	-	-	T	C	-	C	t		C	-
	6654	-	-	T	C	-	T	-		C	-
	7107	NS4B	I113M	A	G	I → M	A	-		g	a
	7577	NS5	D20G	A	G	D → G	G	-		G	-
N5H	889	Env	I12V	A	G	I → V	G	a	1-10223	G	-
	1619	Env	K255R	A	G	K → R	G	a		G	-
	7540	-	-	T	C	-	C	t		C	-
	7577	NS5	D20G	A	G	D → G	G	a		G	-
	7692	-	-	G	A	-	A	g		A	-
	9702	-	-	T	C	-	C	t		C	-
N6H	1601	Env	N249S	A	G	N → S	g	a	1-10236	G	-
	7101	-	-	T	C	-	c	t		C	-
N7H	1058	Env	V68A	T	C	V → A	T	-	1-6210, 6492-7565	c	t
	2186	Env	G445E	G	A	G → E	a	g		G	-
	3567	-	-	A	G	-	A	-		g	a
	4474	NS3	I8V	A	G	I → V	g	a		A	-
	5169	-	-	T	C	-	c	t		T	-
	9879	-	-	T	C	-	c	t		n/a	

N8H	1619	Env	K255R	A	G	K → R	g	a	1-4391, 4633-10236	G	a
	9193	NS5	S559G	A	G	S → G	g	a		g	a
N9H	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	-
N10H	5439	-	-	A	G	-	A	-	14-10228	G	-
N11H	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	-
N14H	2222	Env	I456T	T	C	I → T	C	t	11-10223	C	t
	7107	NS4B	I113M	A	G	I → M	G	a		G	a
N15H	1044	-	-	A	G	-	G	a	16-10236	G	-
	1601	Env	N249S	A	G	N → S	G	a		G	-
	3361	NS1	S343G	A	G	S → G	G	a		G	-
	8937	-	-	T	G	-	G	t		G	-

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%. Italic, grey capital letters indicate major variants that were found in the inoculum but not in the sequence retrieved from hamster liver samples. In the case in which the sequence could not be obtained from the hamster liver samples, the variants letters were replaced by the term “n/a”.

436
437

et al., submitted 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: 98%), but at 6 dpi, severe weight loss was recorded (mean of %IW: 75%). 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. We observed a reduction in mortality rate in all groups infected with re-encoded viruses. For most of them (YFV n1h, n4h, n7h, n9h, n10h, n11h, n14h and n15h), lethality had completely disappeared. However, for YFV n2h and n3h, mortality was drastically reduced but still observed (17 and 20%, respectively) and YFV n5h, n6h and n8h showed mortality rates that were close to that of *Ap7M* (50, 60 and 40% respectively).

Whatever the observed mortality rate, clinical signs of illness were observed with varying degrees of severity amongst groups. At 3 dpi, weight evolution was heterogeneous, with an increase up to 7% in YFV n1h to n4h-infected groups, a slight decrease (between 3 and 6%) in groups infected with YFV n5h to n7h and n11h to n14h, and a greater decrease up to 11% in YFV n8h, n9h and n10h-infected groups. Weight evolution at 3 dpi was not associated to mortality rate, with reported means of %IW ranging between 90 and 107% amongst lethal viruses. At 6 dpi, important weight loss was observed in all groups (up to 45%). Although viruses endowed with important mortality rates (YFV n5h, n6h and n8h) induced important weight loss at 6 dpi (45, 35 and 31%, respectively), some of the non-lethal viruses also did (38 and 33% for YFV n10h and n15h).

At 3 dpi, viral loads in the liver were close to values observed with YF *Ap7M*, ranging from 7.10^8 to 5.10^{10} RNA copies/gram of liver. At 6 dpi, most of them remained close to what was observed with *Ap7M*. However, in some cases, significantly higher (YFV n5h and n6h, 3 and 8.10^9, respectively) or lower (YFV n1h and n11h, 5 and 8.10^7, respectively) viral loads were observed. For most viruses, endpoint viral loads were close to that of *Ap7M* but for YFV n7h,

n9h and n11h, they were significantly lower (between 5.10^6 and 3.10^7). The evolution of viral yields at 3 and 6 dpi was not associated to higher or lower mortality rates. Unsurprisingly, the only association we observed was between survival and endpoint viral loads, with significantly lower viral yields in the liver of surviving hamsters.

Sequence analysis. Some lines of evidence can be drawn by comparing the *in vivo* phenotypes of viruses in the light of the sequences that were retrieved from the infected hamster liver samples (See details in Table 6).

First, the sequence obtained from the liver samples of hamsters infected with the model strain *Ap7M* did not include any mutation. This suggests that, in a virus with unaltered replicative fitness, no adaptative mutations are needed for efficient replication *in vivo*.

For viruses n2h, n7h and n10h, the sequence retrieved from the hamster livers did not include any NS mutation associated to major variants. Virus n2h showed a mild attenuation (low mortality rate and reduction of weight loss at 6 dpi) that may thus be attributed to the 247 simple transitions used for re-encoding. Virus n7h showed a more important attenuation (no mortality, reduction in weight loss, reduction in endpoint viral loads) that can be explained either by the larger extent of re-encoding (326 mutations) and/or by the nature of the mutations used for re-encoding (introduction of UpA dinucleotides). However, for both viruses, only a partial coding sequence could be obtained so these results should be taken cautiously. For virus n10h, the complete coding sequence was obtained from the hamster liver samples so the *in vivo* attenuation of the virus that was observed can be attributed to re-encoding on reasonable grounds. More specifically, it showed the efficiency of the use of transitions, notably on sites potentially involved in secondary structures in providing a stable re-encoded infectious genome exhibiting *in vivo* attenuation.

Three of the re-encoded viruses showed a phenotype close to that of the model strain in terms

of mortality, weight and viral load evolution: n5h, n6h and n8h. All of them exhibited NS mutation(s) associated to major variant(s) in the E (n6h, n5h and n8h) and/or in the NS5 (n5h) proteins coding sequence.

For n6h and n8h, those mutations were not found in the inoculum or, if so, only associated to minor variants, suggesting an adaptation of the re-encoded viruses to *in vivo* replication conditions. Interestingly, some convergence was observed in the adaptation process. The mutation of an asparagine in position 249 within the E protein sequence into a serine (E/N249S) was observed for viruses n1h, n6h and n15h. In the same manner, the E/K255R mutation, was observed in the E protein sequence of viruses n5h and n8h. In the case of viruses n6h and n8h, the NS mutations E/N249S and E/K255R were the only ones to be associated to major variants in the viral sequence and may have allowed the recovery of an *in vivo* phenotype close to that of *Ap7M* virus.

All n5h NS mutations were observed both in the inoculum and in the sequence obtained from liver samples, but two of them were also observed in other viruses, the mutations E/K255R (shared with n8h virus) and NS5/D20G (shared with n4h virus). Although it is not clear whether only one or both of these mutations were determinant in the recovery of an *in vivo* replicative fitness, it is very likely that they took part in the adaptation of those re-encoded viruses to *in vivo* conditions.

The other viruses showed either mild (n3h, n4h and n15h) or strong (n1h, n9h, n11h and n14h) attenuation *in vivo* but for all of them, NS mutations associated to major variants were identified in the viral sequence obtained from liver samples. For variants n1h and n9h, none of the NS mutations could be observed in the inoculum. Therefore, they probably resulted from an adaptation of the re-encoded viruses to *in vivo* conditions during the infection. Nevertheless, as those strains showed an attenuated phenotype when compared to that of *Ap7M*, the molecular adaptation did not allow a full recovery of virulence and thus, of viral fitness *in vivo*.

For viruses n3h and n4h, some of the NS mutations observed in the inoculum were not maintained throughout the passage in hamsters. Hence, they cannot be critical to the

adaptation of re-encoded viruses to the *in vivo* environment. However, new NS mutations that were not observed in the inoculum arose during the infection *in vivo*. The latter probably resulted from an adaptation of the re-encoded viruses to *in vivo* conditions during the infection. All those strains showed a mildly attenuated *in vivo* phenotype, suggesting that, again, the molecular adaptation was not sufficient to allow a full recovery of the *in vivo* replicative fitness. For viruses n11h, n14h and n15h, the NS mutations were also present in the sequence from the inoculum and it is therefore complicated to determine if those mutations were production artefacts or resulted from an adaptation of the virus to culture conditions. An adaptive role can be suggested for the I113M mutation in the NS4B protein, as this NS change was also identified in the sequences of viruses n1h, n4h and n14h. However, neither the I113M nor the other NS mutations allowed a complete recovery of *in vivo* replicative fitness in the case of viruses n11h, n14h and n15h.

Altogether, those results indicate that mutations can arise and stably establish in re-encoded viruses in response to *in vivo* conditions. This was observed for a majority -but not all- of the re-encoded viruses, notably some with as few as 101 re-encoded sites (n1h) and as much as 353 re-encoded sites (n3h). In most cases, those mutations did not allow to fully recover *in vivo* viral replicative fitness. However, convergent adaptation mutations were observed that may have allowed, in some cases, the recovery of an *in vivo* phenotype close to that of the *Ap7M* virus. Observable changes in the biological properties of the viruses (*e.g.* change in tropism, as reported for YF strain 17D), were not reported in any of the viruses that exhibited NS mutations.

Immunisation potential of YF-re-encoded strains: serological investigations. For each of the re-encoded viruses that showed complete loss of mortality *in vivo*, sera were recovered from infected hamsters at 16 dpi during the comparative study of pathogenicity (see above). Sera from both uninfected hamsters and vaccinated humans were included as negative and positive controls, respectively. Sero-neutralisation tests were performed with all

sera, using the YF *Ap7M* virus and BHK21 cells. Neutralising antibodies were detected in both hamsters infected with YF-re-encoded viruses (IC_{50} ranging between 2.10^{-3} and 1.10^{-5}) and in a control group of 17D-vaccinated humans (IC_{50} between 4 and 5.10^{-4}) whilst the negative control group did not show any neutralisation activity against *Ap7M* (see Fig 2).

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 n9h or n11h). A group of 9 hamsters was not immunized before the challenge and a control group of 2 uninfected hamsters was maintained for weight monitoring.

Unexpectedly, one hamster from the YFV n9h 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 n9h and n11h groups, respectively, and a hamster from the non-infected group were euthanized 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 euthanized 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 for neither the immunized nor the 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 n9h and n11h

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 immunized 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 n9h and n11h.

Mutant name	**Mortality**	**Mean viral loads (RNA copies/g of liver)**				**Mean normalized 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%
n9h	0%	2E+07	1E+07	3E+07	1E+07	93%	100%	98%
n11h	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 normalized percentage of initial weight means at day of challenge and 3/6 dpc are given for all groups (*i.e.* immunized and control). Viral loads and normalized 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 including poliovirus, influenza, respiratory syncytial, chikungunya, tick-borne encephalitis and dengue virus (31, 32, 41-43). 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 further characterise the mechanisms that underlie the attenuation process and 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 (41, 42). 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 (48)) and a robust YF hamster model (47). 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 (and deleterious) impact of synonymous substitutions on YFV replicative phenotype both *in vitro* and *in vivo*. As previously described with TBEV (42), the level of re-encoding necessary to observe viral attenuation *in vivo* was significantly lower than that required to obtain observable attenuation *in vitro*. All FII re-encoded strains (300-350 mutations) maintained a replicative phenotype close to that of the parental *Asibi* strain *in vitro*. Only YFV FII/IIIR, 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 the 13 FII re-encoded strains that were tested in hamsters. Whilst the least attenuated viruses showed no significant loss of replicative fitness nor virulence, 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 FII/IIR strain as 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 could result from the adaptation of re-encoded viruses to transfection, *in vitro* culture or *in vivo* conditions. In some cases, we observed the emergence of non-synonymous mutations within the sequence from the viral stock that served as an inoculum for the *in vivo* experiments. For those viruses, the confrontation of the results from *in vivo* experiments to sequencing data was thus limited as NS mutations polluted the determination of the specific impact of silent mutations on the viral replicative phenotype *in vivo*.

In total, we designed thirteen FII-rencoded fragments that were used to produce either *Asibi* or *Ap7M*-derived re-encoded strains. After DNA amplification, transfection and three passages in cell culture, seven of the thirteen *Ap7M*-derived re-encoded strains retained a sequence that did not include any mutation associated to a major variant. After one passage in hamster, three out of these seven viruses arboured a sequence that was similar to the inoculum. All of them showed either mild or significant attenuation *in vivo*. Amongst the four viruses that acquired

non-synonymous mutations associated to major variants *in vivo*, two showed a strong attenuation while the two others had a phenotype close to that of *Ap7M* virus. Five out of the six re-encoded strains that arbored non-synonymous mutations associated to major variants in their inoculum were attenuated *in vivo* while the last one did not.

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 weight of the introduction of simple transitions (Ts) (*i.e.*, with no CpG/UpA introduction and with no modification of potential secondary structures) on the replicative fitness of YFV. Variant n2h, for which sequence remained unchanged along the experiments included 247 Ts. It exhibited a mild attenuation *in vivo* (decreased mortality rate, reduced weight loss at 6 dpi).

Similarly, we observed a mild deleterious effect of the introduction of transitions at sites involved in potential secondary structures. Variant n10h, which sequence also remained stable throughout the study, involved 339 transitions of which 50 were located on potential secondary structures. The strain showed an attenuated phenotype *in vivo* (complete loss of mortality).

We conclude that Ts and Ts located on potential secondary structures are endowed with a low deleterious effect on viral replicative fitness. These constitute novel procedures that are relevant for viral attenuation through re-encoding.

Previous studies have showed that the increase in CpG/UpA dinucleotides may be a crucial component of the attenuation mechanism, through an enhancement of recognition by the host immune system (38, 49). We further confirmed the deleterious effect of the introduction of CpG/UpA dinucleotides on viral replicative fitness. For variant n7h, that involved 326 transitions and 59 additional UpA dinucleotides, no NS mutation was detected in the inoculum nor in the sequence from hamster livers and the phenotype was attenuated *in vivo*. Similarly, the variant n9h had an inoculum sequence identical to the design, involved 325 transitions including 59 supplementary CpG dinucleotides, and showed to be significantly attenuated *in vivo*.

We also found evidence that the use of trinucleotide introduction may impact the *in vivo* phenotype of YFV. Viruses n14h and n15h involved 118 additional TCG trinucleotides. Although both those variants had inoculum sequences that differed from the original design, the attenuated phenotypes they exhibited in hamsters suggest a deleterious effect of TCG trinucleotide introduction on YFV replicative fitness and virulence *in vivo*.

Finally, no definitive conclusion could be drawn regarding the specific impact of the use of simple transversions as viruses n4h and n5h had inoculum sequences that differed from the design and showed either low or no attenuation. Similarly, the impact of ACA trinucleotide introduction on viral replicative fitness remains unclear as n14h and n15h sequences both differed from the design.

The evaluation of the stability of viral re-encoded variants is a crucial element of their characterisation. Here, we did not observe any reversion in YFV re-encoded strains *in vivo* and nearly no reversion *in vitro*. However, for several re-encoded strains, the occurrence of NS mutations in the sequence from 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 (41, 50).

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 optimised and more realistic design would rely on the evenly distribution of synonymous mutations with low fitness weight along the genome associated to *in vivo* serial passage and deep sequencing characterisation 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

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 or BHK21 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.

755

756 The amount of viral RNA was calculated from standard curves using a synthetic RNA transcript
757 (concentrations of 10^7, 10^6, 10^5 and 10^4 copies/µL). Results were normalized using
758 amplification (qRT-PCR) of the housekeeping gene actin (as described by Piorkowski and
759 colleagues (5)). The standard curves generated for all the YFV-specific assays had coefficient
760 of determination values (R2) >0.98 and amplification efficiencies were between 93% and
761 100%.

762

763 **Competition assays**

764

765 WT virus was grown in competition with each one of the 14 re-encoded viruses (YFV n1 to
766 n11, n14, n15 and FII/IIIReenc) using two RNA ratios in triplicate (WT/re-encoded virus: 10/90
767 and 50/50). A global estimated MOI of 0.5 was used for the first inoculation. Viruses from each
768 experiment were then passaged nine times with an incubation time of 3 days. At each passage,
769 the estimated MOI was bottlenecked at approximately 0.6. After each infection (3 dpi), nucleic
770 acids were extracted from the clarified culture supernatant using the QiaCube HT device (see
771 corresponding section) and used for RT-PCR amplification. The detection of each virus was
772 achieved using next generation sequencing (NGS) methods (see corresponding section), by
773 amplifying a 256 nucleotides region and counting virus-specific patterns within the read
774 population.

775

776 ***In vivo* experiments**

777

778 The laboratory hamster model reproduces well the features of human YF but requires the use
779 of adapted strains (7). We recently implemented a hamster model for YFV, based on the use
780 of a strain equivalent to the *Asibi*/hamster p7 strain described by Mac Arthur and colleagues in
781 2003 (8), *Ap7M*. When inoculated to hamsters, this strain induces a lethal viscerotropic disease
782 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 normalized 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 euthanized 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 euthanized 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.

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. King A. AM aLE. Virus Taxonomy. Ninth report of the International Committee on Taxonomy of Viruses.; 2014.
2. 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.
3. 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.
4. Gould EA, Solomon T. Pathogenic flaviviruses. Lancet. 2008;371(9611):500-9.
5. 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.
6. 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.
7. Kuno G, Chang GJ, Tsuchiya KR, Karabatsos N, Cropp CB. Phylogeny of the genus Flavivirus. J Virol. 1998;72(1):73-83.
8. 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).
9. Monath TP, Barrett AD. Pathogenesis and pathophysiology of yellow fever. Adv Virus Res. 2003;60:343-95.
10. Monath TP. Treatment of yellow fever. Antiviral Res. 2008;78(1):116-24.
11. 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.
12. 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.
13. Ellis BR, Barrett AD. The enigma of yellow fever in East Africa. Rev Med Virol. 2008;18(5):331-46.
14. 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.
15. 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.
16. Ahmed QA, Memish ZA. Yellow fever from Angola and Congo: a storm gathers. Trop Doct. 2017;47(2):92-6.
17. 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.
18. 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.
19. Barrett MGS. Yellow fever vaccine. In: Elsevier, editor. Vaccines: Sixth Edition2012. p. 870-968.
20. Monath TP. Review of the risks and benefits of yellow fever vaccination including some new analyses. Expert Rev Vaccines. 2012;11(4):427-48.
21. 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.
22. Monath TP, Vasconcelos PF. Yellow fever. J Clin Virol. 2015;64:160-73.
23. Monath TP. Yellow fever vaccine. Expert Rev Vaccines. 2005;4(4):553-74.
24. Sabin AB, Boulger LR. History of Sabin attenuated poliovirus oral live vaccine strains. Journal of Biological Standardization. 1973;1(2):115-8.
25. Schwarz AJ. Preliminary tests of a highly attenuated measles vaccine. Am J Dis Child. 1962;103:386-9.
26. 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.
27. 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.
28. Buynak EB, Hilleman MR. Live attenuated mumps virus vaccine. 1. Vaccine development. Proc Soc Exp Biol Med. 1966;123(3):768-75.
29. Minor PD. The polio-eradication programme and issues of the end game. J Gen Virol. 2012;93(Pt 3):457-74.
30. 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.
31. 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.
32. 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.
33. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 1985;2(1):13-34.
34. 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.
35. Kudla G, Murray AW, Tollervey D, Plotkin JB. Coding-sequence determinants of gene expression in Escherichia coli. Science. 2009;324(5924):255-8.
36. 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.
37. 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.
38. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. Elife. 2014;3:e04531.
39. Jimenez-Baranda S, Greenbaum B, Manches O, Handler J, Rabadan R, Levine A, et al. Oligonucleotide motifs that disappear during the evolution of influenza virus in humans increase alpha interferon secretion by plasmacytoid dendritic cells. J Virol. 2011;85(8):3893-904.
40. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog. 2008;4(6):e1000079.
41. Nougairede A, De Fabritus L, Aubry F, Gould EA, Holmes EC, de Lamballerie X. Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus in primate and mosquito cells. PLoS Pathog. 2013;9(2):e1003172.
42. de Fabritus L, Nougairede A, Aubry F, Gould EA, de Lamballerie X. Attenuation of tick-

borne encephalitis virus using large-scale random codon re-encoding. PLoS Pathog. 2015;11(3):e1004738.
43. Martinez MA, Jordan-Paiz A, Franco S, Nevot M. Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness. Trends Microbiol. 2016;24(2):134-47.
44. Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH. Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci U S A. 2004;101(10):3480-5.
45. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr., Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009;460(7256):711-6.
46. Thurner C, Witwer C, Hofacker IL, Stadler PF. Conserved RNA secondary structures in Flaviviridae genomes. J Gen Virol. 2004;85(Pt 5):1113-24.
47. 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.
48. Aubry F, Nougairede A, de Fabritus L, Querat G, Gould EA, de Lamballerie X. Single-stranded positive-sense RNA viruses generated in days using infectious subgenomic amplicons. J Gen Virol. 2014;95(Pt 11):2462-7.
49. 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.
50. 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.

Figure Captions

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 (n1h to n11h) 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 (n1 to n11 and n1h to n11h) production, a different subgenomic fragment FII (FIIn1 to FIIn11) was amplified from the corresponding plasmid. ***For YFV reencoded strain FII/IIIReenc production, the subgenomic fragment FIIIR was amplified from the corresponding plasmid and combined to fragments FI (wild-type) and FII (reencoded).

Figure 2. Results of YFV serology at 16 days post-infection. Sero-neutralisation assays 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 negative and positive controls. Results are given as IC50 values evaluated using Graphpad Prism Software. Thick black lines indicate median values and minimum/maximum values are indicated by dashes.

Supporting information

S1 Figure. Dinucleotide introduction bias according to the position in the coding frame within a 35 YFV CDS alignment. Mutations were defined with reference to the *Asibi* strain (Genbank accession number: AY640589).

S2 Figure. ISF-preferred patterns: Comparison of trinucleotides frequencies within the ISF, NKV and DENV groups.

S3 Figure. MBF-preferred patterns: Comparison of trinucleotides frequencies within the ISF, NKV and DENV groups.

S1 Table. Yellow Fever dataset: sequence accession numbers.

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

S3 Table. Potential 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).

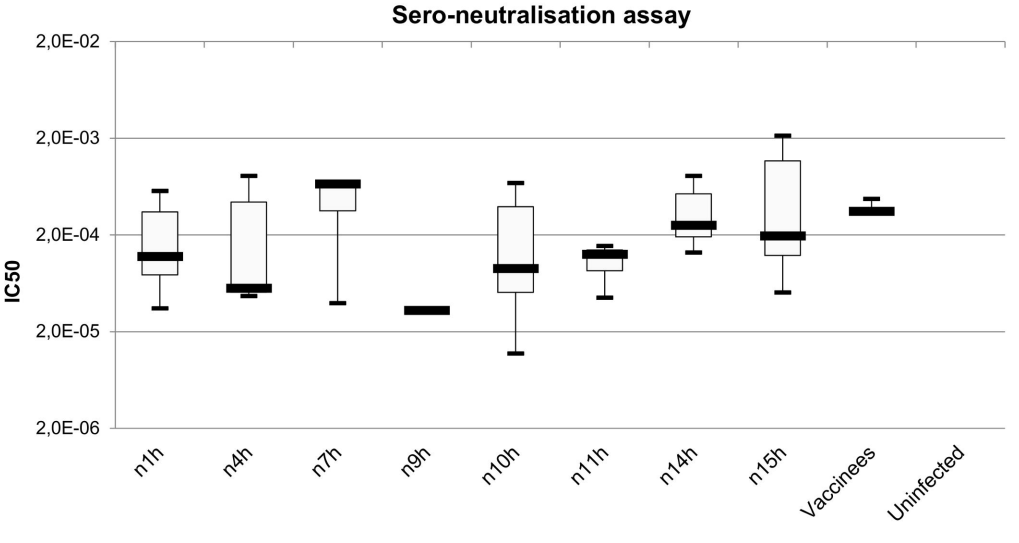
S1 Supplementary Protocols.

S2 Supplementary Results.

Additions to YFV genome for virus production using ISA method.

Aubry F. et al. 2014

Asibi, *Ap7M* and reencoded strains subgenomic fragments


Sero-neutralisation assay
IC50
2,0E-02
2,0E-03
2,0E-04
2,0E-05
2,0E-06
n1h
n4h
n7h
n9h
n10h
n11h
n14h
n15h
Vaccinees
Uninfected