SUPPLEMENTARY DATA FOR

Primed adaptation tolerates extensive structural and size variations of the CRISPR RNA guide in *Haloarcula hispanica*

Luyao Gong[1,2], Ming Li[1,*], Feiyue Cheng[1,2], Dahe Zhao[1], Yihua Chen[1,2] and Hua Xiang[1,2,*]

[1] *State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China.*

[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

Summary of supplementary figures and tables:

Figure S1 3' handle is essential for immunity mediated by the v60 spacer.

Figure S2 CRISPR transcription ended at either of two terminators.

Figure S3 The stem-loop forming potential of truncated 3' handles.

Figure S4 Adaptation to HHPV-2 was primed by the v10-crRNAs carrying different 3' handles.

Figure S5 Verification of the transcription start site (TSS) designed at the very beginning of 5' handle.
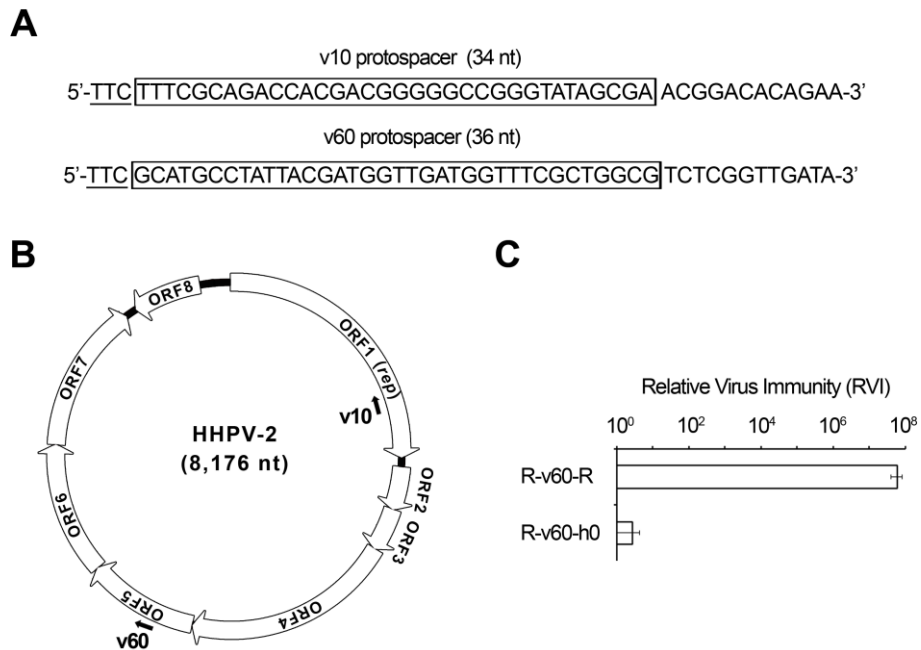
Figure S6 Spacer13 persisted during two spacer loss assays.

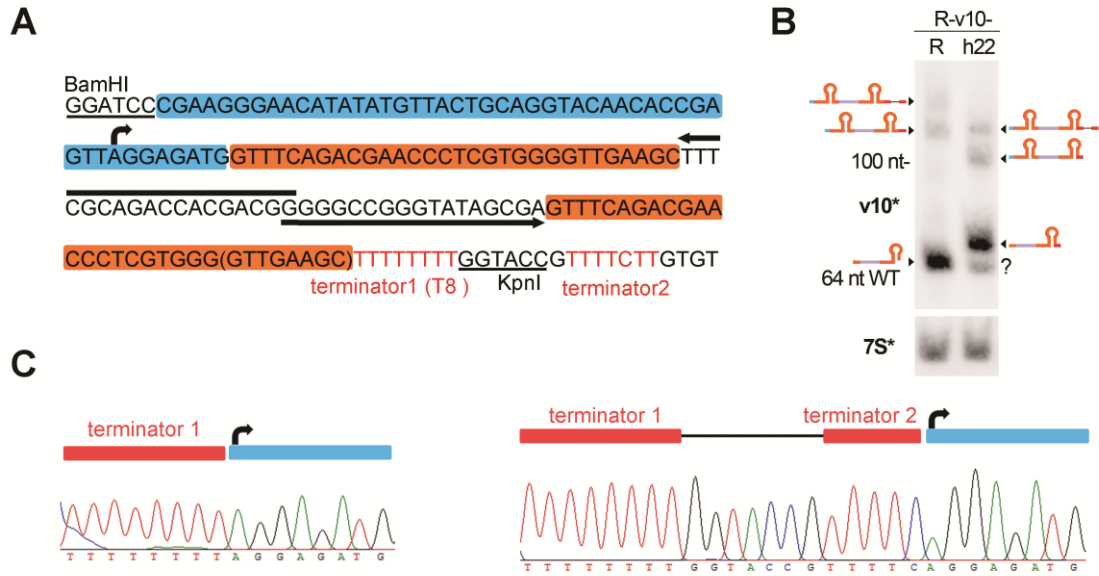**Table S1** Strains and plasmids used in this study.

Table S2 Oligonucleotides used in this study.

Table S3 The virus-derived new spacers during adaptation primed by H6-v10-h0.

SUPPLEMENTARY FIGURES

A

v10 protospacer (34 nt)

5'-TTC TTTCGCAGACCACGACGGGGGCCGGGTATAGCGA ACGGACACAGAA-3'

v60 protospacer (36 nt)

5'-TTC GCATGCCTATTACGATGGTTGATGGTTTCGCTGGCG TCTCGGTTGATA-3'

B

HHPV-2
(8,176 nt)

ORF8
ORF7
ORF6
ORF5
ORF4
ORF3
ORF2
ORF-1 (rep)

v10
v60

C

Relative Virus Immunity (RVI)

10^0 10^2 10^4 10^6 10^8

R-v60-R

R-v60-h0

Figure S1 3' handle is essential for immunity mediated by the v60 spacer. (A) The sequence of the v10 or v60 protospacer is given, with the PAM nucleotides underlined. 12 downstream nucleotides are also shown. **(B)** The location of v10 and v60 protospacers on the viral genome. **(C)** Intense virus immunity was conferred by R-v60-R, which encodes a canonical crRNA, but not conferred by R-v60-h0, which encodes a 3' handle-lacking crRNA. To get the RVI value, the plaque forming unit (PFU) on Sp1d cells containing the empty pWL502 was divided by the PFU on cells containing a CRISPR-expressing plasmid. Error bars indicate the standard deviation (SD) calculated from three independent replicates.

A

BamHI
GGATCC CGAAGGGAACATATATGTTACTGCAGGTACAACACCGA
GTTAGGAGATG GTTTCAGACGAACCCTCGTGGGGTTGAAGC TTT
CGCAGACCACGACGGGGGCCGGGTATAGCGA GTTTCAGACGAA
CCCTCGTGGG(GTTGAAGC) TTTTTTTT GGTACC GTTTTCTT GTGT
terminator1 (T8)   KpnI   terminator2

B

R-v10-
R   h22

100 nt-
v10*
64 nt WT   ?

7S*

C

terminator 1

terminator 1   terminator 2

T T T T T T T A G G A G A T G

T T T T T T T G G T A C C G T T T T C A G G A G A T G

Figure S2 CRISPR transcription ended at either of two terminators. (A) Two terminator sequences (in red) in the R-v10-R (with the nucleotides in brackets) and R-v10-h22 (without the nucleotides in brackets) constructs. The promoter, 5'-UTR, and start codon of *phaR* are shaded blue, and the repeat sequences are shaded orange. The two divergent arrows indicate the locations of primers used for CR-RT-PCR analysis. **(B)** Northern analysis of the RNA products from R-v10-R and R-v10-h22 using a v10-specific probe (v10*). The predicted identity of each RNA product is indicated with cartoons beside the band. The question mark indicates an unexpected RNA product. 7S RNA was probed as the internal control (7S*). **(C)** CR-RT-PCR analysis identifying the two transcription terminators. Two representative sequencing results of the main PCR products are given. Curved arrows indicate transcription initiation.

A
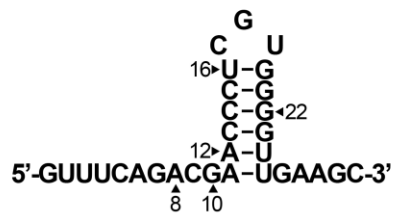
B

Constructs	The ribonucleotides 3' of spacer RNA	MFE (kcal/mol)
R-v10-R	5'-GUUUCAGACGAACCCUCGUGGG-3'	-3.6
R-v10-h22	5'-GUUUCAGACGAACCCUCGUGGG$(U)_8$-3'	-7.6
R-v10-h16	5'-GUUUCAGACGAACCCU$(U)_8$-3'	0
R-v10-h12	5'-GUUUCAGACGAA$(U)_8$-3'	0
R-v10-h10	5'-GUUUCAGACG$(U)_8$-3'	0
R-v10-h8	5'-GUUUCAGA$(U)_8$-3'	0
R-v10-h0	5'-$(U)_8$-3'	0
R-v10-h12m	5'-ACGGACACAGAA$(U)_8$-3'	0

Figure S3 The stem-loop forming potential of truncated 3' handles. (A) The stem-loop forming potential of the WT repeat RNA. Positions relative to the 5' end were given for some nucleotides. **(B)** For each WT or truncated 3' handle, the minimal free energy (MFE) was predicted using the RNAfold web server.
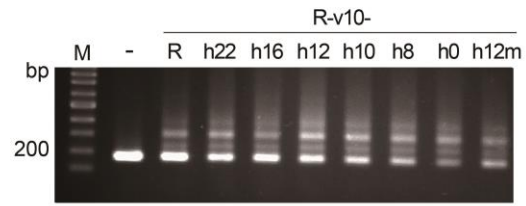
The 2nd PCR analysis; MOI=1:1

Figure S4 Adaptation to HHPV-2 was primed by the v10-crRNAs carrying different 3' handles. The adaptation assay was performed as illustrated in Figure 3A, and included two rounds of infection (MOI=1:1). PCR analysis was performed using primers test_F and test_R.
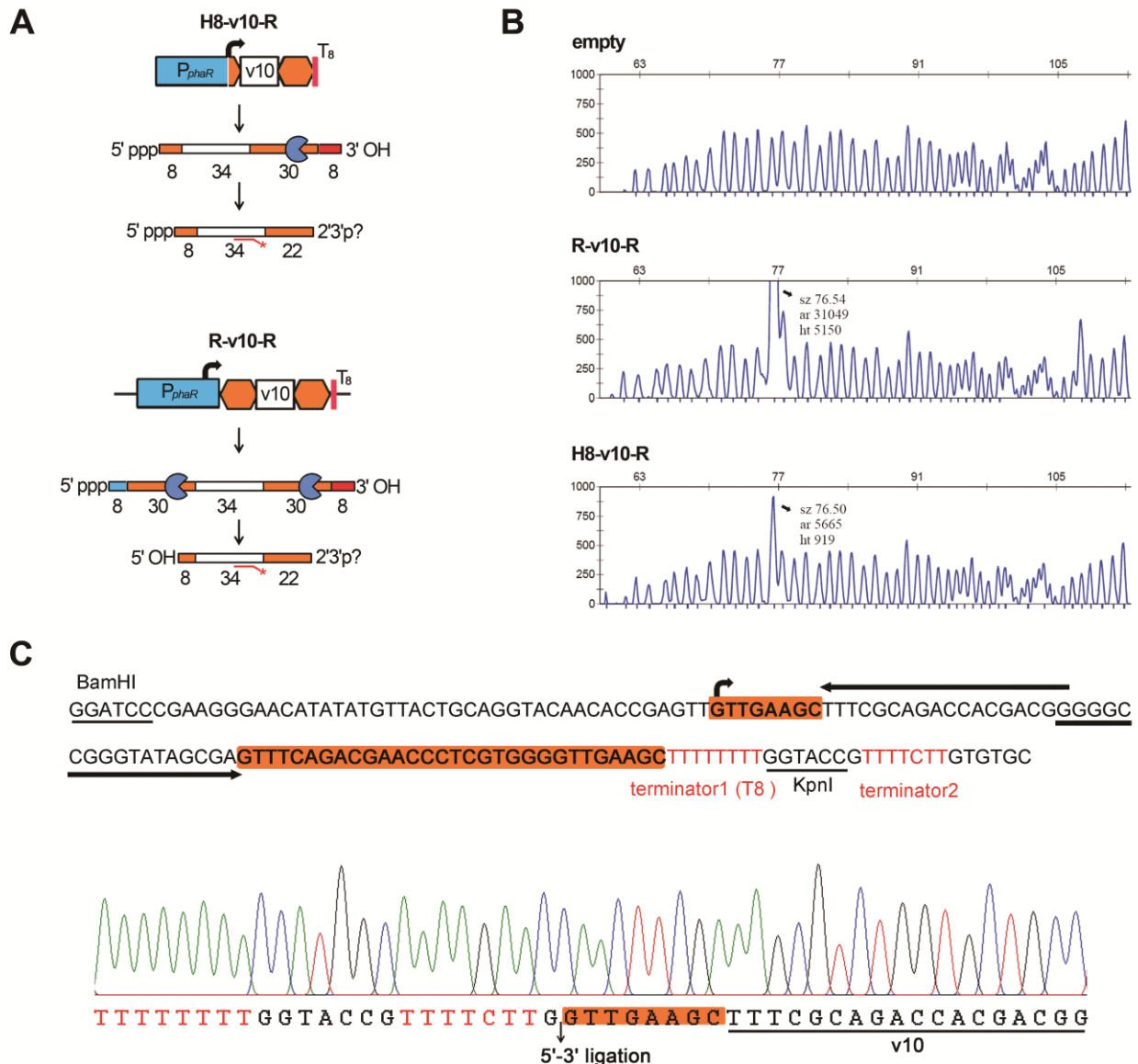
Figure S5 Verification of the transcription start site (TSS) designed at the very beginning of 5' handle. (A) Schematic depicting the structure of the H8-v10-R and R-v10-R constructs, and their RNA products. Their primary transcripts are processed by the Cas6 endonuclease to produce mature cRNAs. The location of the 5' FAM (6-carboxyfluorescein)-labeled extension primer (a red line with an asterisk) is indicated beneath the crRNA molecule. **(B)** Primer extension analysis. H8-v10-R and R-v10-R were separately introduced into Sp1d cells, and total RNA was extracted and subjected to reverse transcription using the primer indicated in panel **A**. The extension products were subjected to fragment analysis using ABI3730xl DNA Analyzer (Thermo Fisher Scientific, MA, USA), and the results were analyzed using GeneMapper 4.1. The consistent location of signal peak indicates the same size of their 5' handle and transcription initiation at the beginning nucleotide of 5' handle in H8-v10-R. **(C)** CR-RT-PCR analysis to confirm the TSS of H8-v10-R. The location of primers is indicated by two long arrows. The curved arrow indicates TSS. The repeat sequence is in bold. The representative sequencing result of PCR products is given.
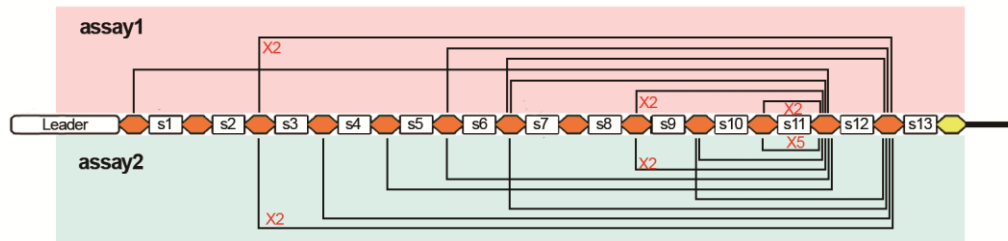
Figure S6 Spacer13 persisted during two spacer loss assays. For each assay, a single DF60 colony was inoculated and serially sub-cultured for more than 20 times, and then subjected to the pTTC11 (a modified pWL502 derivate carrying the PAM sequence 5'-TTC-3' and the protospacer of spacer11) challenge under selection pressure. The CRISPR content of the survivors were analyzed by DNA sequencing. Spacer loss events caused by recombination between the consensus repeats (orange) are indicated (with frequency labeled in red), and no recombination events involving the degenerate repeat (yellow) was observed.

Table S1. Strains and plasmids used in this study

Strains/Plasmids	Description	Source or reference
H. hispanica **strains**		
DF60	*pyrF*-deleted strain of *H. hispanica* ATCC 33960	34
Sp1d (or Δsp1-14)	Modified DF60 with its wild-type CRISPR replaced by a mini-CRISPR containing spacer1 followed by a degenerate repeat	18
Sp13c	Modified DF60 with its wild-type CRISPR replaced by a mini-CRISPR containing spacer13 followed by a consensus repeat	This study
Δ*cas6*	*cas6*-deleted mutant of DF60	This study
Plasmids		
pWL502	7.8 kb; expression vector containing a *pyrF* and its native promoter derived from *Haloferax mediterranei*	35
pTTC1	7.9 kb; modified pWL502 carrying protospacer1 preceded by TTC	21
pTTC11	7.9 kb; modified pWL502 carrying protospacer11 preceded by TTC	This study
pTTC13	7.9 kb; modified pWL502 carrying protospacer13 preceded by TTC	This study
pR-v10-R	7.9 kb; modified pWL502 carrying R-v10-R	This study
pR-v10-h22	7.9 kb; modified pWL502 carrying R-v10-h22	This study
pR-v10-h16	7.9 kb; modified pWL502 carrying R-v10-h16	This study
pR-v10-h12	7.9 kb; modified pWL502 carrying R-v10-h12	This study
pR-v10-h10	7.9 kb; modified pWL502 carrying R-v10-h10	This study
pR-v10-h8	7.9 kb; modified pWL502 carrying R-v10-h8	This study
pR-v10-h0	7.9 kb; modified pWL502 carrying R-v10-h0	This study
pR-v10-h12m	7.9 kb; modified pWL502 carrying R-v10-h12m	This study
pH8-v10-R	7.9 kb; modified pWL502 carrying H8-v10-R	This study
pH8-v10-h22	7.9 kb; modified pWL502 carrying H8-v10-h22	This study
pH8-v10-h0	7.9 kb; modified pWL502 carrying H8-v10-h0	This study
pH7-v10-h0	7.9 kb; modified pWL502 carrying H7-v10-h0	This study
pH6-v10-h0	7.9 kb; modified pWL502 carrying H6-v10-h0	This study
pH5-v10-h0	7.9 kb; modified pWL502 carrying H5-v10-h0	This study
pH4-v10-h0	7.9 kb; modified pWL502 carrying H4-v10-h0	This study
pR-16bp-h22	7.9 kb; modified pWL502 carrying R-16bp-h22	This study
pR-21bp-h22	7.9 kb; modified pWL502 carrying R-21bp-h22	This study
pR-24bp-h22	7.9 kb; modified pWL502 carrying R-24bp-h22	This study
pR-28bp-h22	7.9 kb; modified pWL502 carrying R-28bp-h22	This study
pR-30bp-h22	7.9 kb; modified pWL502 carrying R-30bp-h22	This study
pR-31bp-h22	7.9 kb; modified pWL502 carrying R-31bp-h22	This study
pR-32bp-h22	7.9 kb; modified pWL502 carrying R-32bp-h22	This study
pR-44bp-h22	7.9 kb; modified pWL502 carrying R-44bp-h22	This study

pR-60bp-h22	7.9 kb; modified pWL502 carrying R-60bp-h22	This study
pR-72bp-h22	7.9 kb; modified pWL502 carrying R-72bp-h22	This study
pH8-p1-h22	7.9 kb; modified pWL502 carrying H8-p1-h22	This study
pH8-p1-h0	7.9 kb; modified pWL502 carrying H8-p1-h0	This study
pH7-p1-h0	7.9 kb; modified pWL502 carrying H7-p1-h0	This study
pH6-p1-h0	7.9 kb; modified pWL502 carrying H6-p1-h0	This study
pH5-p1-h0	7.9 kb; modified pWL502 carrying H5-p1-h0	This study
pH6M1-p1-h0	7.9 kb; modified pWL502 carrying H6M1-p1-h0	This study
pH6M2-p1-h0	7.9 kb; modified pWL502 carrying H6M2-p1-h0	This study
pH6M3-p1-h0	7.9 kb; modified pWL502 carrying H6M3-p1-h0	This study
pH6M4-p1-h0	7.9 kb; modified pWL502 carrying H6M4-p1-h0	This study
pH6M5-p1-h0	7.9 kb; modified pWL502 carrying H6M5-p1-h0	This study
pH6M6-p1-h0	7.9 kb; modified pWL502 carrying H6M6-p1-h0	This study
pR-v60-h0	7.9 kb; modified pWL502 carrying R-v60-h0	This study
pR-v60-R	7.9 kb; modified pWL502 carrying R-v60-R	This study
pHAR	4.0 kb; suicide vector containing *pyrF* and its native promoter derived from *Haloferax mediterranei*	34
pDCAS6	5.2 kb; modified pHAR for *cas6* deletion	This study
pSP13C	5.3 kb; modified pHAR for construction of Sp13c	This study

Table S2. Oligonucleotides used in this study

Primer[a]	5'–3' sequence[b]
For strains construction	
Sp13c_UF	CGC<u>GGATCC</u>CTCGCAGACCGGGTACTAT
Sp13c_UR	**TGAAACTCTATCGACCAGCGCG**AAGTAAGACTGTGTGTGCCGAGCTTCAACCCCACGAG
Sp13c_DF	**CGCGCTGGTCGATAGAGTTTCA**GACGAACCCTCGTGGGGTTGAAGCTACGTCAGTGCTC
Sp13c_DR	CGG<u>GGTACC</u>GGGCTTGCCGACTGAA
DCAS6_UF	CGC<u>GGATCC</u>CGAGCAAGGCGCAATAGAT
DCAS6_UR	**TGATTTGTGGTTGGCCACTACACGCACTCGGATGAC**
DCAS6_DF	**GTCATCCGAGTGCGTGTAGTGGCCAACCACAAATCA**
DCAS6_DR	CGG<u>GGTACC</u>CCGGGCCATTGGTACT
For target plasmid construction	
TTC11_F	<u>GATCC</u>TTC CTCGCCGCACTGGCTGGTCGCCAGACCCATCACA<u>GGTAC</u>
TTC11_R	<u>C</u>TGTGATGGGTCTGGCGACCAGCCAGTGCGGCGAGGAA<u>G</u>
TTC13_F	<u>GATCC</u>TTC TCGGCACACACAGTCTTACTTCGCGCTGGTCGATAGA<u>GGTAC</u>
TTC13_R	<u>C</u>TCTATCGACCAGCGCGAAGTAAGACTGTGTGTGCCGAGAA<u>G</u>
For crRNA expression plasmid construction	
Pro_F	CGC<u>GGATCC</u>CGAAGGGAACATATATGT
v10_R1	**CCCGTCGTGGTCTGCGAAAG**CTTCAACCCCACGAGGG
R-v10-h0_R2	CGG<u>GGTACC</u>AAAAAAAATCGCTATACCCGGCC**CCCGTCGTGGTCTGCGAAAG**
R-v10-R_R	CGG<u>GGTACC</u>AAAAAAAAGCTTCAACCCCACGAGGGGTTCGTCTGAAACTCGCTATACCCG
R-v10-h22_R	CGG<u>GGTACC</u>AAAAAAAACCCACGAGGGGTTCGTCTGAAACTCGCTATACCCGGCCCCCG
R-v10-h16_R	CGG<u>GGTACC</u>AAAAAAAAAGGGTTCGTCTGAAACTCGCTATACC
R-v10-h12_R	CGG<u>GGTACC</u>AAAAAAAATTCGTCTGAAACTCGCTATACC
R-v10-h10_R	CGG<u>GGTACC</u>AAAAAAAACGTCTGAAACTCGCTATACC
R-v10-h8_R	CGG<u>GGTACC</u>AAAAAAAATCTGAAACTCGCTATAC
R-v10-h12m_R	CGG<u>GGTACC</u>AAAAAAAATTCTGTGTCCGTTCGCTATACCCGGCC
H8-v10-h0_R1	**TCGCTATACCCGGCCCCCGTCGTGGTCTGCGAAAG**CTTCAACAACTCGGTGTTGTACCT
H7-v10-h0_R1	**GTCGTGGTCTGCGAAAG**CTTCAAAACTCGGTGTTGTACCTG
H6-v10-h0_R1	**GTCGTGGTCTGCGAAAG**CTTCAAACTCGGTGTTGTACCTG
H5-v10-h0_R1	**GTCGTGGTCTGCGAAAG**CTTCAACTCGGTGTTGTACCTG
H4-v10-h0_R1	**GTCGTGGTCTGCGAAAG**CTTAACTCGGTGTTGTACCTG
R-16bp-h22_R	CGG<u>GGTACC</u>AAAAAAAACCCACGAGGGTTCGTCTGAAACGTCGTGGTCTGCG
R-21bp-h22_R	CGG<u>GGTACC</u>AAAAAAAACCCACGAGGGTTCGTCTGAAACCCCCCGTCGTGGTCTGC

R-24bp-h22_R	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACCGGCCCCCGTCGTGG
R-28bp-h22_R	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACTACCCGGCCCCCGTC
R-30bp-h22_R	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACTATACCCGGCCCCCG
R-31bp-h22_R	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACCTATACCCGGCCCCCG
R-32bp-h22_R	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACGCTATACCCGGCCCC
R-44bp-h22_R	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACCTGTGTCCGTTCGCTATACC
R-60bp-h22_R1	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACGTTCCTGATT**CTGTGTCCG**
R-60bp-h22_R2	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACCCTGAAGTTC**CTGATTCTGT**
R-72bp-h22_R1	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAACGTTGCC**CCTGAAGTTCCTG**
R-72bp-h22_R2	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAAC**TCTGCTGTTGCCCCTGAAG**
H8-p1-h22_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTTCAACAACTCGGTGTTGTAC
H8-p1-h22_R2	CGG**GGTACC**AAAAAAAACCCACGAGGGTTCGTCTGAAAC**TTTGAGTCGCTTCGCC**
H8-p1-h0_R2	CGG**GGTACC**AAAAAAAA**TTTGAGTCGCTTCGCC**
H7-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTTCAA**AACTCGGTGTTGT**
H6-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTTCA**AACTCGGTGTTGT**
H5-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTTC**AACTCGGTGTTGT**
H6M1-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCTTCTTCA**AACTCGGTGTTGT**
H6M2-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGATTCA**AACTCGGTGTTGT**
H6M3-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCGTCA**AACTCGGTGTTGT**
H6M4-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTGCA**AACTCGGTGTTGT**
H6M5-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTTAA**AACTCGGTGTTGT**
H6M6-p1-h0_R1	**TTTGAGTCGCTTCGCC**GCCTGTCCCGCAGCCTTGTCGCTTCC**AACTCGGTGTTGT**
v60_R1	**AATAGGCATGC**GCTTCAACCCCACGAGGGTTCGTCTGAAACCATCT

	CCTAACTCGGTGT
R-v60-h0_R2	CGG<u>GGTACC</u>AAAAAAAACGCCAGCGAAACCATCAACCATCGT**AATA** **GGCATGC**
R-v60-R	CGG<u>GGTACC</u>AAAAAAAAGCTTCAACCCCACGAGGGTTCGTCTGAAA CCGCCAGCGAAAC

Adaptation analysis

test_F	CGC<u>GGATCC</u>TCGGTTTCCGTCGAC
test_R	CGG<u>GGTACC</u>CGGAGGCGATGACTGATG

Northern analysis

7S-probe-biotin	GGGGGCGTCCGGTCTGA
s1-probe-biotin	TGACTGATGGCAGAAAGCAA
s13-probe-biotin	CGACCAGCGCGAAGTAA
v10-probe-biotin	CCGTCGTGGTCTGCGAAA

CR-RT-PCR analysis

CR-RT_F	GGGGCCGGGTATAGCGA
CR-RT_R	CCGTCGTGGTCTGCGAAA

Primer extension analysis

FAM-v10	TTCTGCTGTTGCCCCTGAAGTTCCTGATTCTGTGTCCGTTCGCTATA CCCGGCCCCCGT

[a] Forward and reverse primers are respectively indicated with letters F and R. UF/UR and DF/DR are primers respectively for the upstream and downstream fragments, which were subsequently linked by bridge PCR reactions. R1 and R2 primers contain overlapping sequences that are designed for overlap extension PCR.

[b] Restriction endonuclease cut sites are underlined. Complementary sequences within UR and DF primers (for construction of CRISPR variant strains), and overlapping sequences within R1 and R2 primers (for construction of crRNA-expressing plasmids), are indicated in bold. Restriction endonuclease cut sites of BamHI or KpnI are underlined. The mutated nucleotides in primers for 5' handle single mutation are in red. The designed PAM nucleotides are shown in box.

Table S3. The virus-derived new spacers during adaptation primed by H6-v10-h0

New spacer sequences	Positions[a]	Strand[b]	PAM
CCTTCGGCAATAGCTTCAGACGCAACGTCAGTATC	1715-1681	-	TTC
CGGGTCTGTACCTGTCGAATCGGTCCCTGACGAAA	72-106	+	TTC
GACGGGCTATCATCACCATTCTCAGATGCAAACACA	4566-4531	-	TTC
ACTATCGAAACACAGGCGTTTACGATTATTCGGTC	5866-5900	+	TTC
AGGGAACGACTCATCTTCATTCCACGATCCGAAGA	1165-1131	-	TTC
GCCCAGCTCAACAACGCCGTAATCACCAGCATCAGGG	5764-5728	-	TTC
ATGCTACCCGCTCCGGATCAACCGTACGGCCAGCCA	2885-2850	-	TTC
GAGATCATCAAGCGCAGCGCCCGCGATCTCGCAAA	1510-1476	-	TTC
AGGATGAGCTGGGCGAGTTGCAAGATGATGAGTTC	6017-6051	+	TTC
GATGCTCGGACTAATCGGCGTGAAGTTGCTGTGCAG	6834-6869	+	TTC

[a]The positions of protospacers (from which a spacer was derived) on the HHPV-2 genome.

[b]'+' and '-' correspond respectively to the coding strand and the template strand of the *rep* gene (ORF1).