SUPPLEMENTAL FIGURES

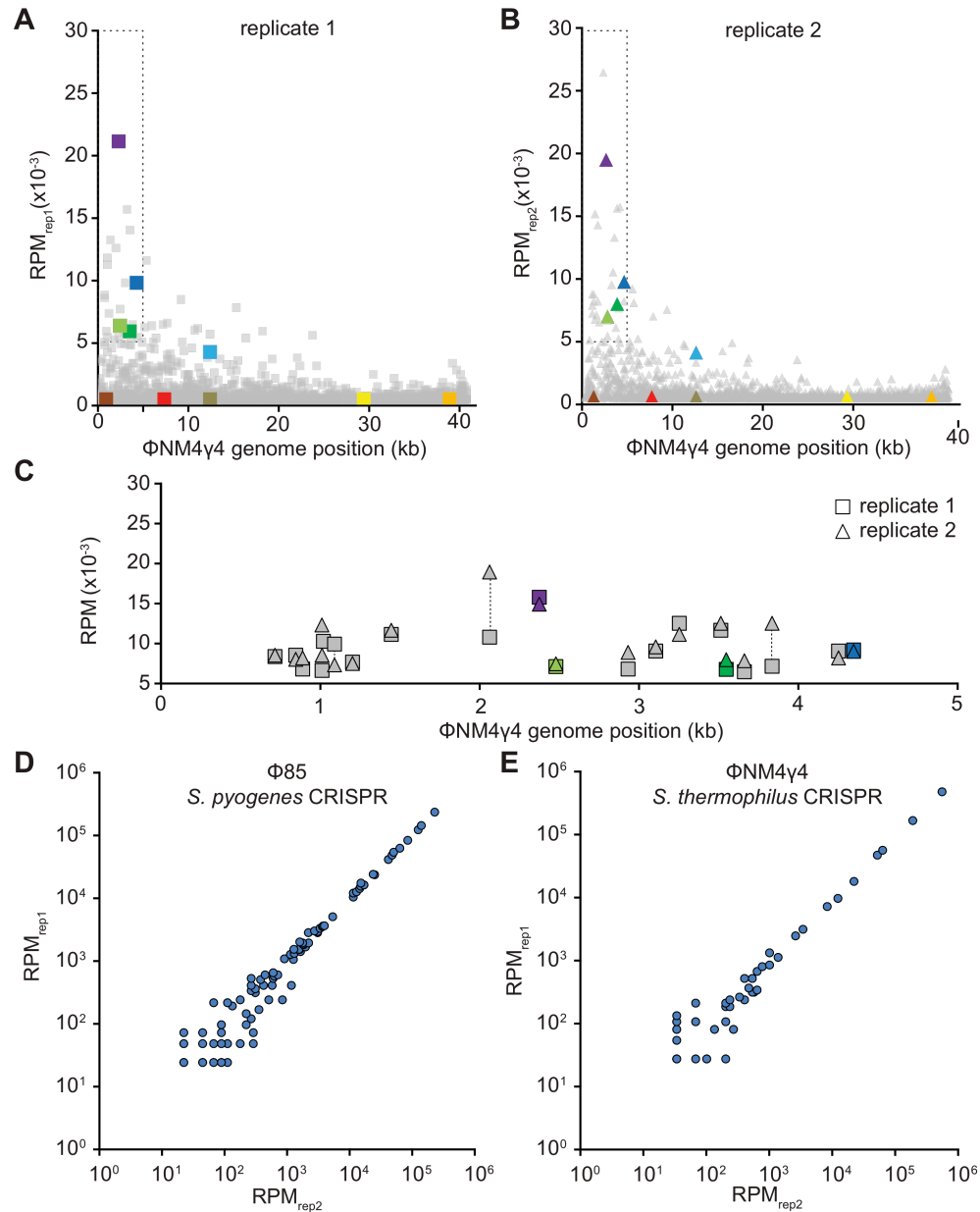**Figure S1. Biased sampling of phage DNA protospacers is a feature of other bacteriophages and type II CRISPR-Cas systems. Related to Fig. 1.** (**A**), (**B**) Abundance (in reads per million, RPM) of φNM4γ4 viral sequences incorporated as spacers into the CRISPR array, mapped against location on the phage genome, in duplicate (raw data for Figure 1B and 1C). (**C**) Overlap of data in (**A**) and (**B**), zoomed on the first 5 kb of the viral genome. Only spacers with RPM > 5,000 are shown. (**D**) RPM values of spacers sampled in two replicates during infection with lytic phage φ85 of cells harboring the *Streptococcus pyogenes* CRISPR-Cas system. (**E**) RPM values of spacers sampled in two replicates during infection with φNM4γ4 of cells harboring the *Streptococcus thermophilus* type II-A CRISPR3 system.
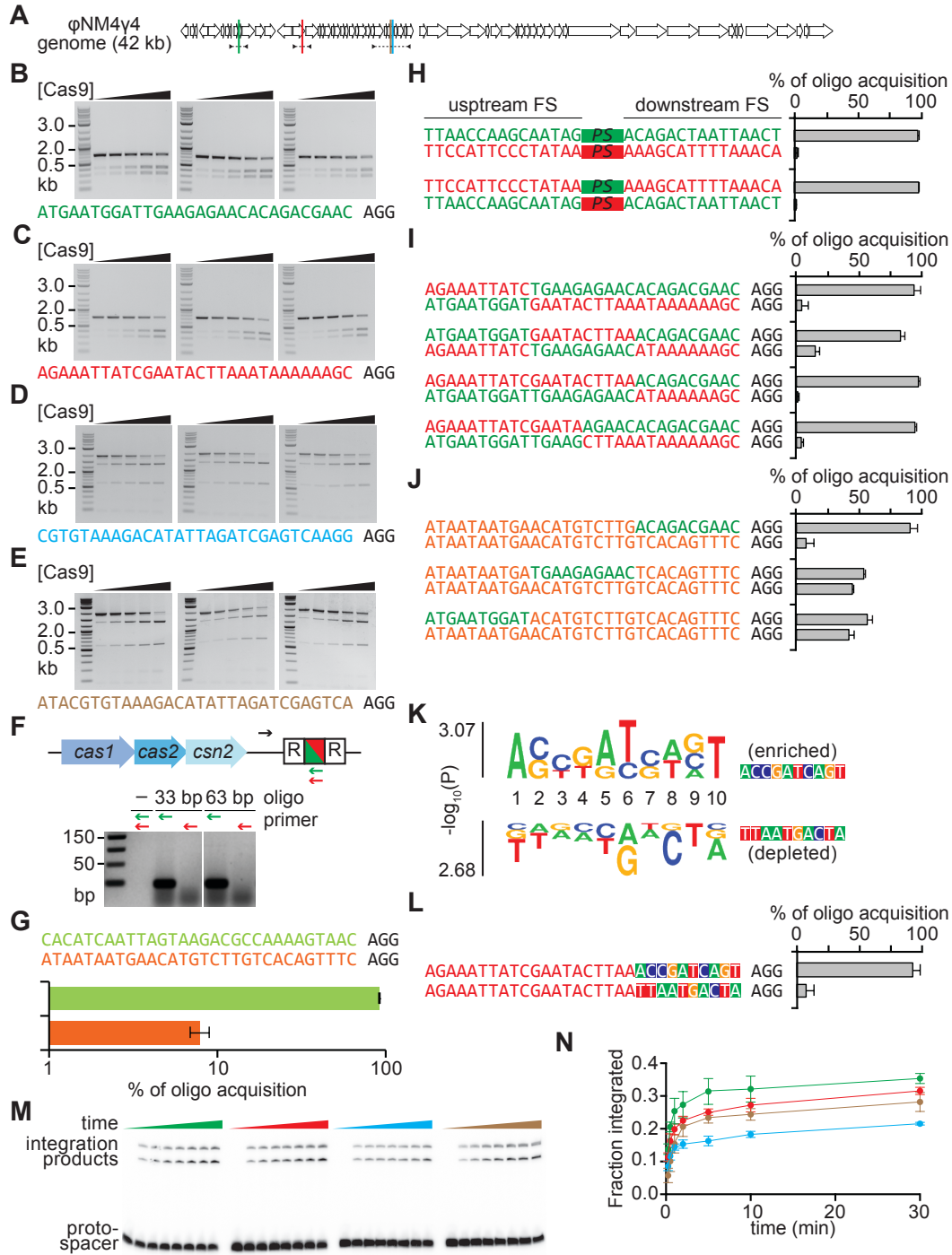
A φNM4γ4 genome (42 kb)

B [Cas9]
3.0
2.0
0.5
kb
ATGAATGGATTGAAGAGAACACAGACGAAC AGG

C [Cas9]
3.0
2.0
0.5
kb
AGAAATTATCGAATACTTAAATAAAAAAGC AGG

D [Cas9]
3.0
2.0
0.5
kb
CGTGTAAAGACATATTAGATCGAGTCAAGG AGG

E [Cas9]
3.0
2.0
0.5
kb
ATACGTGTAAAGACATATTAGATCGAGTCA AGG

F cas1 cas2 csn2 R R
— 33 bp 63 bp oligo
primer
150
50
bp

G
CACATCAATTAGTAAGACGCCAAAAGTAAC AGG
ATAATAATGAACATGTCTTGTCACAGTTTC AGG
1 10 100
% of oligo acquisition

M
time
integration products
proto-spacer

H
usptream FS   downstream FS
% of oligo acquisition
0   50   100
TTAACCAAGCAATAG *PS* ACAGACTAATTAACT
TTCCATTCCCTATAA *PS* AAAGCATTTTAAACA

TTCCATTCCCTATAA *PS* AAAGCATTTTAAACA
TTAACCAAGCAATAG *PS* ACAGACTAATTAACT

I
% of oligo acquisition
0   50   100
AGAAATTATCTGAAGAGAACACAGACGAAC AGG
ATGAATGGATGAATACTTAAATAAAAAAGC AGG

ATGAATGGATGAATACTTAAACAGACGAAC AGG
AGAAATTATCTGAAGAGAACATAAAAAAGC AGG

AGAAATTATCGAATACTTAAACAGACGAAC AGG
ATGAATGGATTGAAGAGAACATAAAAAAGC AGG

AGAAATTATCGAATAAGAACACAGACGAAC AGG
ATGAATGGATTGAAGCTTAAATAAAAAAGC AGG

J
% of oligo acquisition
0   50   100
ATAATAATGAACATGTCTTGACAGACGAAC AGG
ATAATAATGAACATGTCTTGTCACAGTTTC AGG

ATAATAATGATGAAGAGAACTCACAGTTTC AGG
ATAATAATGAACATGTCTTGTCACAGTTTC AGG

ATGAATGGATACATGTCTTGTCACAGTTTC AGG
ATAATAATGAACATGTCTTGTCACAGTTTC AGG

K
3.07
-log₁₀(P)
A C G A T C A G T (enriched)
ACCGATCAGT
1 2 3 4 5 6 7 8 9 10
(depleted)
TTAATGACTA
2.68

L
% of oligo acquisition
0   50   100
AGAAATTATCGAATACTTAAACCGATCAGT AGG
AGAAATTATCGAATACTTAATTAATGACTA AGG

N
0.4
0.3
0.2
0.1
0.0
Fraction integrated
0   10   20   30
time (min)

Figure S2. High and low abundance spacers have different acquisition rates but similar interference efficiencies. Related to Fig. 2. (**A**) Location on the phage genome of the protospacers specified by the spacers colored dark green, red, light blue and tan in Figure 2A. Arrowheads show the target region amplified to generate substrates for in vitro cleavage assays. (**B-E**) In vitro cleavage of amplified phage targets (10 nM) incubated for 5 minutes with various concentrations of Cas9 (6.26, 12.5, 25, 50 and 100 nM) loaded with sgRNAs specified by the four spacers shown in Figure

2A; in triplicate. Cleavage products were separated by agarose gel electrophoresis and stained with ethidium bromide. (**F**) Qualitative PCR assay to assess the integration of low abundance (Figure 1A, red) and high abundance (Figure 1A, dark green) spacers after electroporation of the corresponding dsDNA oligonucleotides mixed in a 1:1 molar ratio. Two dsDNA oligonucleotides were tested: 33 bp (protospacer + PAM sequences) or 63 bp (15 bp upstream + protospacer + PAM + 15 bp downstream sequences). Black arrow, forward primer; green and red arrows; reverse primers that specifically anneal on the integrated spacers. (**G**) Relative acquisition rates (%) of spacers following electroporation of a pairs of high/low abundance dsDNA oligonucleotides (light green/orange spacers in Figure 1C). (**H**) Relative acquisition rates (%) of spacers following electroporation of 2 pairs of dsDNA oligonucleotides with swapped 15-nt upstream and downstream flanking sequences of the dark green and red targets. PS, protospacer. (**I**). Relative acquisition rates (%) of spacers following electroporation of 4 pairs of dsDNA oligonucleotides with swapped 10- or 15-nt spacer sequences of the dark green and red targets. (**J**). Relative acquisition rates (%) of spacers following electroporation of 3 pairs of dsDNA oligonucleotides. In each pair, one of the dsDNA oligonucleotides has the low abundance, orange spacer sequence shown in (**G**) and is co-electroporated with another containing a different PAM-distal, central or PAM-proximal, 10-nt sequence from the dark green spacer (Figure 2A).with swapped 10- or 15-nt spacer sequences of the dark green and red targets. (**K**) Unweighted probability logo of the top 1% acquired spacer sequences generated using kpLogo (showing only 10 bp, PAM-proximal sequence). Nucleotides shown on top were enriched, while the ones shown on the bottom were depleted in the spacers used to create the logo. Enriched or depleted consensus sequences are shown to the right. (**L**). Relative acquisition rates (%) of spacers following electroporation of two dsDNA oligonucleotides, each containing the PAM-distal and central regions of the low abundance, red spacer (Figure 2A) and the 10-nt, PAM-proximal, enriched or depleted sequences identified in (**K**). (**M**) Radiolabeled dsDNA oligonucleotides with the low and high abundant spacer sequences shown in Figure 2A, containing 4-nt 3' overhangs, were incubated with a plasmid carrying the type II-A CRISPR array and purified *S. pyogenes* Cas1-Cas2 complex for increasing amounts of time (0.25, 0.5, 1, 2, 5, 10, and 30 minutes). The products of the reaction were separated by ethidium bromide-stained agarose gel electrophoresis. (**N**) shown in (**M**).
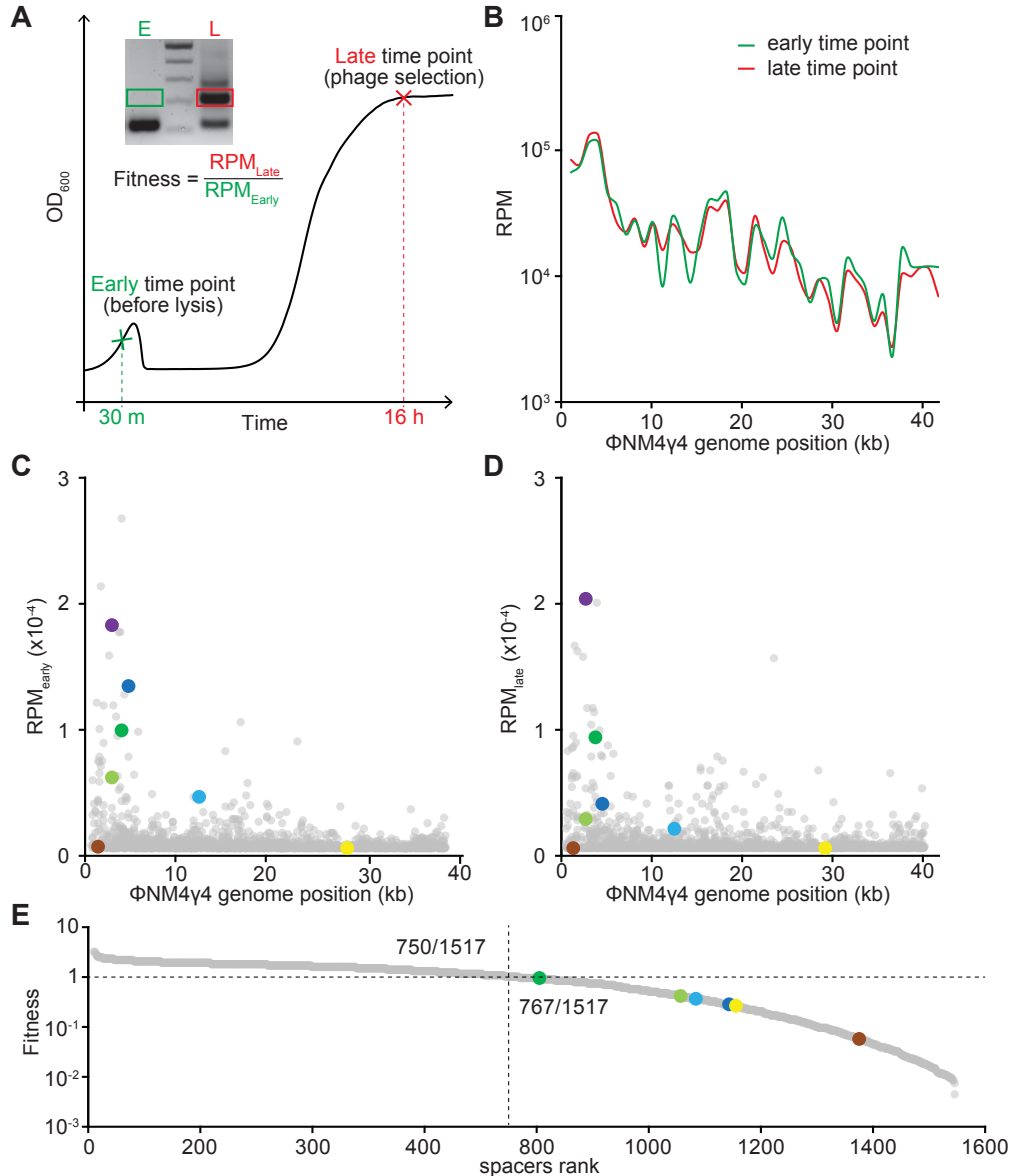
Figure S3. Spacer distribution due to initial acquisition is not perturbed over the course of phage infection. Related to Fig. 3. (**A**) Diagram of assay used to measure the effects of targeting efficiency on spacer abundance. Typically, when challenged by the lytic phage φNM4γ4, staphylococci harboring the *S. pyogenes* type II-A CRISPR-Cas system continue growing for ~ 45 minutes (measured by the OD_{600} of the culture) and then succumb to infection. The small proportion of cells that acquire new spacers survive and enable the regrowth of the culture. DNA was prepared from cells isolated after 30 minutes (early time point) and 16 hours (late time point) of infection. Inset: amplification of the CRISPR array from the extracted DNA at each time point. Boxes show the PCR product corresponding to the expanded CRISPR locus, which is barely visible in the early sample due to the low proportion of staphylococci that acquire new spacers. (**B**) Average abundance (RPM per 1-kb bins) of φNM4γ4 viral sequences incorporated as spacers into the CRISPR array, mapped against location on the phage

genome, in the early and late time point libraries (red and green traces). (**C**) Abundance (in reads per million, RPM) of φNM4γ4 viral sequences incorporated as spacers into the CRISPR array, mapped against location on the phage genome in the early time point. (**D**) Same as (**C**), for late time point. (**E**) Spacers ranked by decreasing fitness (ratio of abundance in late time point divided by abundance in early time point).
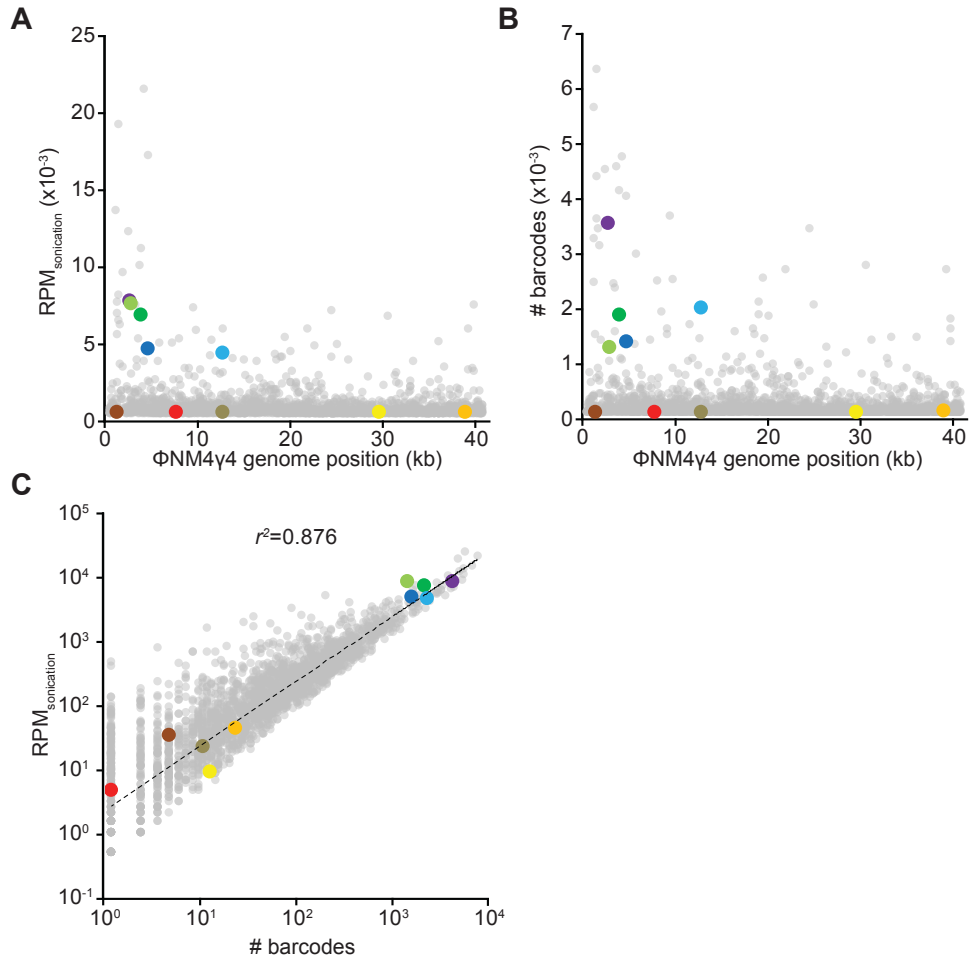
Figure S4. Oversampling of phage protospacers is due to higher rates of acquisition. Related to Fig. 4. (**A**) Abundance (RPM) of spacers incorporated into the CRISPR array, mapped against location on the phage genome, following electroporation of sheared phage DNA. (**B**) Same as (**A**), but plotting the number of different barcodes for each acquired spacer sequence. (**C**) Abundance of individual spacers following electroporation of sonicated phage DNA plotted against barcodes as a measure of the number of times each spacer was acquired.

SUPPLEMENTAL TABLES

Table S1. Oligonucleotides used in this study. Related to STAR methods.

Name	Sequence
B153	CTCGTACAGTGAACCTTTTTCACC
H180	TCTGGTAGAAAAGATATCCTACGAG
H186	NNNNNGTCCAAAATTTTTTAGACAAAAATAGTC
H293	GCAAAAATGGATAAGAAATACTCAATAGGC
H294	TATTGAGTATTTCTTATCCATTTTTGCCTCC
H360	NNNNNGCTTAGCTGAGACAAATAGTGCG
H366	NNNNNGCTAAAACAGCATAGCTCTAAAACG
H366	NNNNNGCTAAAACAGCATAGCTCTAAAACG
H370	NNNNNGACAGGGGCTTTTCAAGACTG
H371	NNNNNGACGAAGAAATCAACCAGCGC
H372	NNNNNACTAGGGGCTTTTCAAGACTG
H378	CAGGGGCTTTTCAAGACTGNNNNNNNNNNNGAGACAAATAGTGCG
H379	CAGTCTTGAAAAGCCCCTG
H433	AAACAAACAGTGACAGAAACTATTGAGTACGAGGG
H434	AAAACCCTCGTACTCAATAGTTTCTGTCACTGTTT
H435	AAACAGAAAACAGTGACAGAAACTATTGAGTACGG
H436	AAAACCGTACTCAATAGTTTCTGTCACTGTTTTCT
H521	GCGGCCTCTAATACGACTCACTATAGGGCATATTAGATCGAGTCAAGGGTTTTAGAGCTAGAAATAGCA
H522	GCGGCCTCTAATACGACTCACTATAGGGAGACATATTAGATCGAGTCAGTTTTAGAGCTAGAAATAGCA
H555	NNNNNTCACTCGTACAGTGAACCTTTTTCACC
H557	CAGAATCCACGAGATCTGTGCCAGTTCGTAATGTCTGG
H558	TACGAACTGGCACAGATCTCGTGGATTCTGTGATTTGG
H559	CCTTCAGGTTATGACCCGTTTGTTGAACTAATGGGTGC
H612	ATGAATGGATTGAAGAGAACACAGACGAACAGG
H613	CCTGTTCGTCTGTGTTCTCTTCAATCCATTCAT
H614	TTAACCAAGCAATAGATGAATGGATTGAAGAGAACACAGACGAACAGGACAGACTAATTAACT
H615	AGTTAATTAGTCTGTCCTGTTCGTCTGTGTTCTCTTCAATCCATTCATCTATTGCTTGGTTAA
H616	AGAAATTATCGAATACTTAAATAAAAAAGCAGG
H617	CCTGCTTTTTTATTTAAGTATTCGATAATTTCT
H618	TTCCATTCCCTATAAAGAAATTATCGAATACTTAAATAAAAAAGCAGGAAAGCATTTTAAACA
H619	TGTTTAAAATGCTTTCCTGCTTTTTTATTTAAGTATTCGATAATTTCTTTATAGGGAATGGAA
H620	ATGAATGGATTGAAGCTTAAATAAAAAAGCAGG
H621	CCTGCTTTTTTATTTAAGCTTCAATCCATTCAT
H622	AGAAATTATCGAATAAGAACACAGACGAACAGG
H623	CCTGTTCGTCTGTGTTCTTATTCGATAATTTCT
H624	ATGAATGGATTGAAGAGAACATAAAAAAGCAGG
H625	CCTGCTTTTTTATGTTCTCTTCAATCCATTCAT

H626	ATGAATGGATGAATACTTAAACAGACGAACAGG
H627	CCTGTTCGTCTGTTTAAGTATTCATCCATTCAT
H628	ATGAATGGATGAATACTTAAATAAAAAAGCAGG
H629	CCTGCTTTTTTATTTAAGTATTCATCCATTCAT
H630	AGAAATTATCTGAAGAGAACACAGACGAACAGG
H631	CCTGTTCGTCTGTGTTCTCTTCAGATAATTTCT
H632	AGAAATTATCTGAAGAGAACATAAAAAAGCAGG
H633	CCTGCTTTTTTATGTTCTCTTCAGATAATTTCT
H634	AGAAATTATCGAATACTTAAACAGACGAACAGG
H635	CCTGTTCGTCTGTTTAAGTATTCGATAATTTCT
H636	TTCCATTCCCTATAAATGAATGGATTGAAGAGAACACAGACGAACAGGAAAGCATTTTAAACA
H637	TGTTTAAAATGCTTTCCTGTTCGTCTGTGTTCTCTTCAATCCATTCATTTATAGGGAATGGAA
H638	TTAACCAAGCAATAGAGAAATTATCGAATACTTAAATAAAAAAGCAGGACAGACTAATTAACT
H639	AGTTAATTAGTCTGTCCTGCTTTTTTATTTAAGTATTCGATAATTTCTCTATTGCTTGGTTAA
H641	AAACATGAATGGATTGAAGAGAACACAGACGAACG
H642	AAAACGTTCGTCTGTGTTCTCTTCAATCCATTCAT
H643	AAACAGAAATTATCGAATACTTAAATAAAAAAGCG
H644	AAAACGCTTTTTTATTTAAGTATTCGATAATTTCT
H645	ATGAATGGATTGAAGAGAACACAGACGAACATT
H646	AATGTTCGTCTGTGTTCTCTTCAATCCATTCAT
H647	TTAACCAAGCAATAGATGAATGGATTGAAGAGAACACAGACGAACATTACAGACTAATTAACT
H648	AGTTAATTAGTCTGTAATGTTCGTCTGTGTTCTCTTCAATCCATTCATCTATTGCTTGGTTAA
H649	AGAAATTATCGAATACTTAAATAAAAAAGCATT
H650	AATGCTTTTTTATTTAAGTATTCGATAATTTCT
H651	TTCCATTCCCTATAAAGAAATTATCGAATACTTAAATAAAAAAGCATTAAAGCATTTTAAACA
H652	TGTTTAAAATGCTTTAATGCTTTTTTATTTAAGTATTCGATAATTTCTTTATAGGGAATGGAA
H653	TTCCATTCCCTATAAAGAAATTATCGAATACTTAAATAAAAAAGCAGGGGGAGGGGTTTAAACA
H654	TGTTTAAACCCCTCCCCTGCTTTTTTATTTAAGTATTCGATAATTTCTTTATAGGGAATGGAA
H655	CACATCAATTAGTAAGACGCCAAAAGTAACAGG
H656	CCTGTTACTTTTGGCGTCTTACTAATTGATGTG
H657	ATAATAATGAACATGTCTTGTCACAGTTTCAGG
H658	CCTGAAACTGTGACAAGACATGTTCATTATTAT
H661	AGAAATTATCGAATACTTAATCACAGTTTCAGG
H668	ATAATAATGAACATGTCTTGACAGACGAACAGG
H669	CCTGTTCGTCTGTCAAGACATGTTCATTATTAT
H670	ATAATAATGATGAAGAGAACTCACAGTTTCAGG
H671	CCTGAAACTGTGAGTTCTCTTCATCATTATTAT
H672	ATGAATGGATACATGTCTTGTCACAGTTTCAGG
H673	CCTGAAACTGTGACAAGACATGTATCCATTCAT
H662	CCTGAAACTGTGATTAAGTATTCGATAATTTCT
H690	AAAGGATACGTGTAAAGACATATTAGATCGAGTCAAGGAGGTTTTG

H691	CAAAACCTCCTTGACTCGATCTAATATGTCTTTACACGTATCCTTT
H694	CCTCTAATACGACTCACTATAGGTGAAGAGAACACAGACGAACGTTTAAGAGCTATGC
H695	CCTCTAATACGACTCACTATAGGAATACTTAAATAAAAAAGCGTTTAAGAGCTATGC
H700	TATAAAGAAATTATCGAATACTTAAACCGATCAGTAGGAAAGC
H701	GCTTTCCTACTGATCGGTTTAAGTATTCGATAATTTCTTTATA
H702	TATAAAGAAATTATCGAATACTTAATTAATGACTAAGGAAAGC
H703	GCTTTCCTTAGTCATTAATTAAGTATTCGATAATTTCTTTATA
JM126	ATGAAGATTATTTCTTAATAACTAAAAATATGGTATAATTAATACCAGCAGTCGGATACC
JM127	GGTATCCGACTGCTGGTATTAATTATACCATATTTTTAGTTATTAAGAAATAATCTTCAT