SUPPLEMENTARY DATA FOR

The spacer size of I-B CRISPR is modulated by the terminal sequence of the protospacer

Ming Li¹, Luyao Gong¹,², Dahe Zhao¹, Jian Zhou¹, Hua Xiang¹,²,*
¹State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China.
²College of Life Science, University of Chinese Academy of Sciences, Beijing, China.

Summary of supplementary figures and tables:

Figure S1. New spacers are non-functional in SgPA.
Figure S2. The size heterogeneity of 604 spacers from haloarchaeal I-B CRISPRs
Table S1. Strains and plasmids used in this study
Table S2. Oligonucleotides used in this study
Table S3. Self DNA-derived spacers
Table S4. Atypical PAMs that were caused by slippage of the protospacer 5’-end cutting
Table S5. Slippage at the PAM-end usually caused an altered spacer size
Table S6. Conservation of the third 3’-end nucleotide within the prevalent spacer sequence of a group
SUPPLEMENTARY FIGURES

Figure S1. New spacers are non-functional in SgPA. Plaque assay were performed using the DF60PA and SgPA strains, as well as their derivates where a virus-targeting spacer (s-1) had been engineered into their a-CRISPR. For the plaque assay, three replicates were performed for each strain to get an average PFU (plaque forming unit) value.
Figure S2 The size heterogeneity of 604 spacers from haloarchaeal I-B CRISPRs. (A) A total number of 604 spacer sequences were collected from the CRISPRdb database (1) and manually curated. For example, the CRISPR orientation was checked based on the leader sequence and the repeat conservation. (B) The spacer size distribution which generally fits a normal curve (the dotted line). The ratio of spacers of a specific size was labeled above columns.
**SUPPLEMENTARY TABLES**

Table S1. Strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strains/Plasmids</th>
<th>Description</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H. hispanica strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DF60</td>
<td>pyrF-deleted strain of <em>H. hispanica</em> ATCC 33960</td>
<td>(2)</td>
</tr>
<tr>
<td>DF60P</td>
<td>DF60 with its wild-type CRISPR substituted by p-CRISPR</td>
<td>(3)</td>
</tr>
<tr>
<td>DF60PA</td>
<td>DF60P transformed by the integrative plasmid pCR-A</td>
<td>(3)</td>
</tr>
<tr>
<td>SgPA</td>
<td>DF60P transformed by the integrative plasmid pSg-A</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pHAR</td>
<td>4.0 kb; suicide vector containing pyrF and its native promoter</td>
<td>(2)</td>
</tr>
<tr>
<td>pHAR-in</td>
<td>4.4 kb; modified pHAR with a 460-bp chromosomal sequence</td>
<td>(3)</td>
</tr>
<tr>
<td>pCR-A</td>
<td>4.6 kb; modified pHAR-in with an a-CRISPR structure containing the complete 105-bp leader and a single repeat</td>
<td>(3)</td>
</tr>
<tr>
<td>pSg-A</td>
<td>4.6 kb; modified pCR-A with the last ten repeat nucleotides mutated</td>
<td>This study</td>
</tr>
<tr>
<td>pWL502</td>
<td>7.9 kb; expression vector containing a pyrF and its native promoter derived from <em>Haloferax mediterranei</em></td>
<td>(2)</td>
</tr>
<tr>
<td>pVS</td>
<td>8.3 kb; modified pWL502 carrying a viral sequence</td>
<td>(4)</td>
</tr>
<tr>
<td>p7908mut</td>
<td>8.3 kb; modified pVS with the 33rd and 34th nucleotides (C33 and T34) downstream of the viral PAM+7908 mutated</td>
<td>This study</td>
</tr>
<tr>
<td>p7981mut</td>
<td>8.3 kb; modified pVS with the 35th nucleotide (C35) downstream of the viral PAM+7981 mutated</td>
<td>This study</td>
</tr>
</tbody>
</table>
## Table S2. Oligonucleotides used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>5'–3' sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For CRISPR expansion detection</strong></td>
<td></td>
</tr>
<tr>
<td>Exp-Fa</td>
<td>CGGGGTCATCCTCGTCTC</td>
</tr>
<tr>
<td>Exp-Ra</td>
<td>CCGCGAACAAGCTCAACAG</td>
</tr>
<tr>
<td>Exp-7908-F</td>
<td>**[TTGGTCTGTGTCCTACGTTCTGAAA]**CTGAATCGTGTCGTCTGTTC</td>
</tr>
<tr>
<td>Exp-7981-F</td>
<td>**[TTGGTCTGTGTCCTACGTTCTGAAA]**AGTATGGTGACGCGGTTG</td>
</tr>
<tr>
<td>s2-primer</td>
<td>CGGGGTACCAGATAGAGCGGATTCCTCGAC</td>
</tr>
<tr>
<td><strong>For SgPA construction</strong></td>
<td></td>
</tr>
<tr>
<td>SgPA-F</td>
<td>CGCGGATCCCTCGTGTTCCGTGCAC</td>
</tr>
<tr>
<td>SgPA-R</td>
<td>CGGGGTACC<strong>TAGGACCAAA</strong>ACGAGGTTCGTC</td>
</tr>
<tr>
<td><strong>For pVS construction and point mutation</strong></td>
<td></td>
</tr>
<tr>
<td>VS-F</td>
<td>CGCGGATCCCTCGTGTTCCGTGTCAC</td>
</tr>
<tr>
<td>VS-R</td>
<td>CGGGGTACCACGCTCAACCATCTAC</td>
</tr>
<tr>
<td>7908M-F</td>
<td>CGGGGATCGTCTCCATGACGACG</td>
</tr>
<tr>
<td>7908M-R</td>
<td>GTCATTCG4ACCACCTCC</td>
</tr>
<tr>
<td>7981M-F</td>
<td>GAATGTCTGCTGTCGCGGTTGTC</td>
</tr>
<tr>
<td>7981M-R</td>
<td>GACCAGACACGCAACACATC</td>
</tr>
</tbody>
</table>

*Forward and reverse primers are indicated with letters F and R, respectively.

*Designed restriction sequences are underlined. mutated nucleotides are shown in bold and italic. A consistent sequence (framed) was designed on the 5'-end of the Exp-7908-F and Exp-7981-F primers.
<table>
<thead>
<tr>
<th>spacer_ID(^a)</th>
<th>spacer_seq(^b) (5’–3’)</th>
<th>derivation(^c)</th>
<th>PAM (5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;contig_00404_s-1</td>
<td>A(_{CTCCGTTTAGCAGCCTATGGGCGTATATCGGGCATG} _GAC</td>
<td>chr2</td>
<td>GAC</td>
</tr>
<tr>
<td>&gt;contig_00994_s-1</td>
<td>CGCTGGACAGTCCCGACGTGGTAGAAGCCCTG</td>
<td>pHH400</td>
<td>TTC</td>
</tr>
<tr>
<td>&gt;contig_10352_s-1</td>
<td>GATTCGTGGTCACTGGGGGTCAGTACCGGCTCA</td>
<td>chr2</td>
<td>TTC</td>
</tr>
<tr>
<td>&gt;contig_13737_s-1</td>
<td>A(_{CTCCGTTAGCAGCCTATGGGCGTATATCGGGCATG} _GAC</td>
<td>chr2</td>
<td>TTC</td>
</tr>
<tr>
<td>&gt;contig_17175_s-1</td>
<td>CAGG(_{CCTCCAGTCTAGCAGCCTATGGGCGTATATCGGGCATG} _GAC</td>
<td>chr1</td>
<td>TCT</td>
</tr>
<tr>
<td>&gt;contig_24068_s-1</td>
<td>TGAGAAACCCTAGCGCAAGGCGGCTGGTGGGA</td>
<td>chr2</td>
<td>TTC</td>
</tr>
</tbody>
</table>

\(^a\)The spacer ID follows the identifier of high-throughput sequencing reads, and ‘s-1’ indicates this new spacer was the most leader-distal one (first acquired).

\(^b\)The spacer nucleotides different from the protospacer sequence are underlined.

\(^c\)The protospacer of the self-derived spacers located on the main chromosome (chr1), the mini-chromosome (chr2) or the mega-plasmid (pHH400).
Table S4. Atypical PAMs that were caused by slippage of the protospacer 5’-end cutting

<table>
<thead>
<tr>
<th>atypical PAM</th>
<th>PAM_events</th>
<th>context&lt;sup&gt;a&lt;/sup&gt;</th>
<th>slip_type</th>
<th>slip_events</th>
<th>slip_ratio&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>28</td>
<td>ATTTC</td>
<td>-1 nt</td>
<td>21</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATTTC</td>
<td>-2 nt</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>CTT</td>
<td>9</td>
<td>CTTTC</td>
<td>-1 nt</td>
<td>6</td>
<td>88.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CTTTC</td>
<td>-2 nt</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GTT</td>
<td>24</td>
<td>GTTC</td>
<td>-1 nt</td>
<td>22</td>
<td>95.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTTTC</td>
<td>-2 nt</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>TTT</td>
<td>202</td>
<td>TTC</td>
<td>-1 nt</td>
<td>194</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTTTC</td>
<td>-2 nt</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

**NTT_total** 263 -- -- 261 99.2%

| AGT          | 6          | AGTTC               | -2 nt     | 6           | 100.0%                |
| CGT          | 40         | CGTTC               | -2 nt     | 38          | 95.0%                 |
| GAT          | 4          | GATTC               | -2 nt     | 3           | 100.0%                |

**NNT_total** 53 -- -- 50 94.3%

| TCA          | 295        | TTCA                | +1 nt     | 294         | 99.7%                 |
| TCC          | 130        | TTCC                | +1 nt     | 130         | 100.0%                |
| TCG          | 378        | TTCC                | +1 nt     | 378         | 100.0%                |
| TCT          | 133        | TTCT                | +1 nt     | 131         | 98.5%                 |

**TCN_total** 936 -- -- 933 99.7%

| CAA          | 3          | TTCAA               | +2 nt     | 2           | 66.7%                 |
| CAG          | 5          | TTCAG               | +2 nt     | 4           | 80.0%                 |
| CTA          | 2          | TTCCTA              | +3 nt     | 1           | 50.0%                 |

**CNN_total** 10 -- -- 7 70.0%

**Total** -- 1262 -- -- 1251 99.1%

<sup>a</sup> The first protospacer-nucleotide(s) immediately downstream of the atypical PAM are shown in frame, while its upstream sequences are underlined.

<sup>b</sup> The ratios in this column were calculated by dividing the number in the ‘slip_events’ column by the number in the ‘PAM_events’ column.
Table S5. Slippage at the PAM-end usually caused an altered spacer size

<table>
<thead>
<tr>
<th>Group_id</th>
<th>Slip</th>
<th>Prevalent size (bp)</th>
<th>Freq. of the prevalent size</th>
<th>Average size (bp)</th>
<th>St.dev. (bp)</th>
<th>Size alteration per slip(a)</th>
<th>Prevalent spacer sequence(c)(5′–3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAM+3867</td>
<td>0 nt</td>
<td>36</td>
<td>37.2%(120/323)</td>
<td>36.03</td>
<td>1.03</td>
<td>1.2007</td>
<td>CGATGGGGAATGCTGTTGATGCTGTGGACTGAGC</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>35</td>
<td>86.7%(26/30)</td>
<td>34.83</td>
<td>0.46</td>
<td>GATGGGGAATGCTGTTGATGCTGTGGACTGAGC</td>
<td></td>
</tr>
<tr>
<td>PAM+5940</td>
<td>0 nt</td>
<td>36</td>
<td>45.0%(77/171)</td>
<td>35.18</td>
<td>1.03</td>
<td>0.5314</td>
<td>GAAAGGGCACCGTCGTTGACGGTGAGTCGGCTG</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>35</td>
<td>64.4%(38/59)</td>
<td>34.64</td>
<td>0.80</td>
<td>AAGAGGGCACCGTCGTTGACGGTGAGTCGGCTG</td>
<td></td>
</tr>
<tr>
<td>PAM+6450</td>
<td>0 nt</td>
<td>36</td>
<td>73.8%(1051/1424)</td>
<td>35.66</td>
<td>0.70</td>
<td>1.1712</td>
<td>GAAAAACGTGATAGCTGCACTGCTGAGGCCTA</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>35</td>
<td>66.0%(124/188)</td>
<td>34.48</td>
<td>0.78</td>
<td></td>
<td>AAAACGTGATAGCTGCACTGCTGAGGCCTA</td>
</tr>
<tr>
<td>PAM+6719</td>
<td>0 nt</td>
<td>34</td>
<td>33.0%(88/267)</td>
<td>35.22</td>
<td>1.14</td>
<td>1.0769</td>
<td>TCTGCTGCGTGTGATGCTGTTACCTCTTG</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>33</td>
<td>44.2%(19/43)</td>
<td>34.14</td>
<td>1.10</td>
<td></td>
<td>CTGCTGCGTGTGATGCTGTTACCTCTTG</td>
</tr>
<tr>
<td>PAM+7061</td>
<td>0 nt</td>
<td>35</td>
<td>57.5%(206/358)</td>
<td>35.16</td>
<td>0.74</td>
<td>1.0616</td>
<td>AGTGGTGATGCTGACTGCGCTGAGGAACAGACCTG</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>34</td>
<td>57.1%(12/21)</td>
<td>34.10</td>
<td>0.64</td>
<td></td>
<td>GTGTTGAGTTGACTGCGTCAAGGAACAGACCTG</td>
</tr>
<tr>
<td>PAM+6422</td>
<td>-2 nt</td>
<td>38</td>
<td>97.4%(37/38)</td>
<td>37.95</td>
<td>0.32</td>
<td>1.0322</td>
<td>TCCTTGTGATAGCTGCACTGCGCTGAGGAAAAACG</td>
</tr>
<tr>
<td></td>
<td>0 nt</td>
<td>36</td>
<td>69.5%(196/282)</td>
<td>35.88</td>
<td>0.67</td>
<td></td>
<td>GCTGATACGTCGCTGAGTTGCAAGGAAAAACG</td>
</tr>
<tr>
<td>PAM-2843</td>
<td>0 nt</td>
<td>36</td>
<td>56.82%(50/88)</td>
<td>35.53</td>
<td>0.77</td>
<td>1.0341</td>
<td>AGACGAGGGCGAACGCAGCGGAGGAAGTACCCATCGA</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>35</td>
<td>62.5%(50/80)</td>
<td>34.50</td>
<td>0.87</td>
<td></td>
<td>GACGAGGGCGAACGCAGCGGAGGAAGTACCCATCGA</td>
</tr>
<tr>
<td>PAM-3439</td>
<td>-1 nt</td>
<td>36</td>
<td>55.17%(16/29)</td>
<td>35.79</td>
<td>0.73</td>
<td>0.5559</td>
<td>CTTTATGATACGTCGCTGACGGTTCAGGCCAAGG</td>
</tr>
<tr>
<td></td>
<td>0 nt</td>
<td>35</td>
<td>37.55%(95/253)</td>
<td>35.24</td>
<td>1.03</td>
<td></td>
<td>CTTTATGATACGTCGCTGACGGTTCAGGCCAAGG</td>
</tr>
<tr>
<td>PAM-674</td>
<td>-1 nt</td>
<td>35</td>
<td>62.5%(20/32)</td>
<td>35.28</td>
<td>0.99</td>
<td>0.28</td>
<td>CCGCGATCTCCGCGAGCGTGATCAAAGGCGGGCCGA</td>
</tr>
<tr>
<td></td>
<td>0 nt</td>
<td>34</td>
<td>40.59%(384/946)</td>
<td>35.00</td>
<td>1.26</td>
<td>1.25</td>
<td>CCGCGATCTCGCGAGCGTGATCAAAGGCGGGCCGA</td>
</tr>
<tr>
<td></td>
<td>+1 nt</td>
<td>33</td>
<td>62.5%(10/16)</td>
<td>33.75</td>
<td>1.06</td>
<td></td>
<td>GCGATCTCCCCAGCGTGAATCAAAGGCGGGCCGA</td>
</tr>
</tbody>
</table>

\(a\)Each spacer group indicates the collection of spacers with a specific 5′-TTC-3′ as its bona fide PAM. For example, the minus-strand TTC sequence at positions 674-676 is the PAM for spacers from the PAM-674 group.

\(b\)This number is calculated by dividing the average size discrepancy between the ‘TTC’ and the ‘slip’ categories by the slippage value (1 or 2 nt). Note that every 1 nt slippage altered the average spacer size by 0.919±0.336 bp on average.

\(c\)The third 3′-end nucleotide (which preferred to be a cytosine) is underlined for each spacer sequence.
**Table S6.** Conservation of the third 3’-end nucleotide within the prevalent spacer sequence of a group

<table>
<thead>
<tr>
<th>Group_id&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Prevalent spacer size (bp)</th>
<th>Prevalent spacer sequence&lt;sup&gt;b&lt;/sup&gt;(5’–3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unimodal-distribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAM+5285</td>
<td>35</td>
<td>GCTATATGGCTGGCCGTACGGTTGTGATCCGGAGGGG</td>
</tr>
<tr>
<td>PAM+6872</td>
<td>35</td>
<td>ACTCCGTTAGCAAAACGCTTTTGGCAACTGAAACGT</td>
</tr>
<tr>
<td>PAM+7781</td>
<td>35</td>
<td>GTGGGACGCTTGATCCTGTGATGAGACACAGCTG</td>
</tr>
<tr>
<td>PAM+7747</td>
<td>36</td>
<td>AAGCATAAAGGCAATGACCCGCTTTTGAGTTGACGCT</td>
</tr>
<tr>
<td>PAM-738</td>
<td>36</td>
<td>TCACTCCATACGTAGCGCAGCACGACGCAACGAGCAG</td>
</tr>
<tr>
<td>PAM-797</td>
<td>37</td>
<td>CGGTATGTGGCTCACCTTGAGTCCACCAGGACGAGC</td>
</tr>
<tr>
<td><strong>Bimodal-distribution</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAM+2133</td>
<td>37</td>
<td>CCGCTACCTCCGTCGCTACCTGCTGCCTGACGCTGCT</td>
</tr>
<tr>
<td>PAM+5190</td>
<td>34, 37</td>
<td>GCTGATGCTGATGAGGCAACGCTTTGAGTGAGGCTA</td>
</tr>
<tr>
<td>PAM+6803</td>
<td>37</td>
<td>GTGCTGATGCTGATGAGGCAACGCACTTGATGCTGCT</td>
</tr>
<tr>
<td>PAM-483</td>
<td>36</td>
<td>AGCCCGCTGCTGCTTTTATCTTGATACGAGCCAGCTC</td>
</tr>
<tr>
<td>PAM-1163</td>
<td>34</td>
<td>CTTTTTCGTTATGACGACCCACATGATGACGCCCAGC</td>
</tr>
<tr>
<td>PAM-6245</td>
<td>34</td>
<td>GAATCACCAGTACATCTCTTGATACACCAGCCAGC</td>
</tr>
</tbody>
</table>

<sup>a</sup>Only the spacer groups listed in Figure 3 of the main text are included in this table.

<sup>b</sup>The third last nucleotide of each spacer sequence is underlined.
REFERENCES


