

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

A thermostable Cas9 with increased lifetime in human plasma

Lucas B. Harrington¹, David Paez-Espino², Janice S. Chen¹, Enbo Ma¹, Brett T. Staahl¹,
Nikos C. Kyrpides² & Jennifer Doudna^{1, 3-6}

¹Department of Molecular and Cell Biology, University of California, Berkeley, California, 94720, USA.

²Department of Energy, Joint Genome Institute, Walnut Creek, California 94598, USA

³Department of Chemistry, University of California, Berkeley, California, 94720, USA.

⁴Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.

⁵Innovative Genomics Institute, University of California, Berkeley, California 94720, USA.

⁶MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

20 **CRISPR-Cas9 is a powerful technology that has enabled genome editing in a wide**
21 **range of species. However, the currently developed Cas9 homologs all originate**
22 **from mesophilic bacteria, making them susceptible to proteolytic degradation and**
23 **unsuitable for applications requiring function at elevated temperatures. Here, we**
24 **show that the Cas9 protein from the thermophilic bacterium *Geobacillus***
25 ***stearothermophilus* (GeoCas9) catalyzes RNA-guided DNA cleavage over a wide**
26 **temperature range and has an enhanced protein lifetime in human plasma.**
27 **GeoCas9 is active at temperatures up to 70°C, compared to 45°C for**
28 ***Streptococcus pyogenes* Cas9 (SpyCas9), which greatly expands the temperature**
29 **range for CRISPR-Cas9 applications. By comparing features of two closely related**
30 ***Geobacillus* homologs, we created a variant of GeoCas9 that doubles the DNA**
31 **target sequences that can be recognized by this system. We also found that**
32 **GeoCas9 is an effective tool for editing mammalian genomes when delivered as a**
33 **ribonucleoprotein (RNP) complex. Together with an increased lifetime in human**
34 **plasma, the thermostable GeoCas9 provides the foundation for improved RNP**
35 **delivery *in vivo* and expands the temperature range of CRISPR-Cas9.**
36

37 The use of CRISPR-Cas9 has rapidly transformed the ability to edit and
38 modulate the genomes of a wide range of organisms¹. This technology, derived from
39 adaptive immune systems found in thousands of bacterial species, relies on RNA-guided
40 recognition and cleavage of invasive viral and plasmid DNA². The Cas9 proteins from
41 these species differ widely in their size and cleavage activities³⁻⁵. Despite the abundance
42 and diversity of these systems, the vast majority of applications have employed the first
43 Cas9 homolog developed from *Streptococcus pyogenes* (SpyCas9)⁶. In addition to
44 SpyCas9, several other Cas9 and related proteins have also been shown to edit
45 mammalian genomes with varying efficiencies^{5,7-10}. While these proteins together
46 provide a robust set of tools, they all originate from mesophilic hosts, making them
47 unsuitable for applications requiring activity at higher temperatures or extended protein
48 stability.

49 This temperature restriction is particularly limiting for genome editing in obligate
50 thermophiles¹¹. Recent efforts using SpyCas9 to edit a facultative thermophile have
51 been possible by reducing the temperature within the organism¹². While effective, this
52 approach is not feasible for obligate thermophiles, and requires additional steps for
53 moderate thermophiles. This is especially important for metabolic engineering for which
54 thermophilic bacteria present enticing hosts for chemical synthesis due to decreased risk
55 of contamination, continuous recovery of volatile products and the ability to conduct
56 reactions that are thermodynamically unfavorable in mesophilic hosts¹³. Developing a
57 thermostable Cas9 system will enable facile genome editing in thermophilic organisms
58 using technology that is currently restricted to mesophiles.

59 CRISPR-Cas9 has also emerged as a potential treatment for genetic diseases¹⁴.
60 A promising method for the delivery of Cas9 into patients or organisms is the injection of
61 preassembled Cas9 ribonucleoprotein complexes (RNP) into the target tissue or
62 bloodstream¹⁵. One major challenge to this approach is that Cas9 must be stable

63 enough to survive degradation by proteases and RNases in the blood or target tissue for
64 efficient delivery. Limited protein lifetime will require delivery of higher doses of Cas9 into
65 the patient or result in poor editing efficacy. In contrast, delivering a Cas9 with improved
66 stability could greatly enhance genome-editing efficiency *in vivo*.

67 To address these challenges, we tested the thermostable Cas9 protein from
68 *Geobacillus stearothermophilus* (GeoCas9). We find that GeoCas9 maintains activity
69 over a wide temperature range. By harnessing the natural sequence variation of
70 GeoCas9 from closely related species, we engineered a protospacer adjacent motif
71 (PAM)-binding variant that recognizes additional PAM sequences and thereby doubles
72 the number of targets accessible to this system. We also engineered a highly efficient
73 single-guide RNA (sgRNA) based on RNAseq data from the native organism and show
74 that GeoCas9 can efficiently edit genomic DNA in mammalian cells. The functional
75 temperature range of GeoCas9 complements that of previously developed Cas9
76 systems, greatly expanding the utility and stability of Cas9 for both *in vitro* cleavage and
77 genome editing applications.

78

79 **Results**

80 ***Identification of thermostable Cas9 homologs***

81 Although thousands of *cas9* homologs have been sequenced, there have been
82 no functionally validated Cas9 orthologs from archaea¹⁶, restricting our search for a
83 thermophilic Cas9 to thermophilic bacteria. We examined all the isolates in Integral
84 Microbial Genomes database (IMG) from a thermophilic environment that contained a
85 Cas9-like protein¹⁷ (hits to a TIGRfam model 01865 for Csn1-like or 03031 for a Csx12-
86 like). From this analysis, the Cas9 from *Geobacillus stearothermophilus* (*G. st.*; formerly
87 *Bacillus stearothermophilus*)¹⁸ stood out because it was full-length and its sequence is
88 shorter than that of the average Cas9. Most importantly, this candidate is found in an =

89 organism that can grow at a reported temperature range of 30°C–75°C (optimal at
90 55°C). A BLASTn search for proteins similar to GeoCas9 revealed several nearly
91 identical homologs (from 93.19–99.91% identity over the full length) in six other
92 *Geobacillus* species (ED Table 2) and 92.55% Identity over the full length in
93 *Effusibacillus pohliae* DSM22757. *G. st.* has been a proven source of enzymes for
94 thermophilic molecular cloning applications¹⁹, thermostable proteases²⁰ and enzymes for
95 metabolic engineering²¹. Moreover, the wide temperature range that *G. st.* occupies²²
96 suggested that the Cas9 from this species (GeoCas9) might able to maintain activity at
97 both mesophilic and thermophilic temperatures (Fig. 1a). Notably, GeoCas9 is
98 considerably smaller than SpyCas9 (GeoCas9, 1,087 amino acids; SpyCas9, 1,368
99 amino acids). A homology model of GeoCas9 based on available Cas9 crystal structures
100 along with sequence alignments revealed that the small size of GeoCas9 is largely the
101 result of a reduced REC lobe, as is the case with other compact Cas9 homologs from
102 *Streptococcus aureus* Cas9 (SauCas9) and *Actinomyces naeslundii* Cas9 (AnaCas9)
103 (Fig. 1b, c; Supplementary Fig. 1).

104 We purified GeoCas9 and performed initial thermostability tests using differential
105 scanning calorimetry (DSC), which showed that in the absence of RNA or DNA,
106 GeoCas9 has a denaturation temperature about 20°C higher than SpyCas9 (Fig. 1d).
107 Moreover, GeoCas9 denatures at 15°C higher than the slightly thermophilic
108 *Streptococcus thermophilus* CRISPR III Cas9 (SthCas9) (Fig. 1e). Given these results,
109 we selected GeoCas9 as a candidate for further development and optimization.

110

111 **GeoCas9 PAM identification and engineering**

112 CRISPR systems have evolved a preference for a protospacer adjacent motif
113 (PAM) to avoid self-targeting of the host genome^{23,24}. These PAM sequences are
114 divergent among Cas9 homologs and bacteriophage DNA targets are often mutated in

115 this region to escape cleavage by Cas9²⁵. To identify the PAM for GeoCas9, we first
116 searched for naturally targeted viral and plasmid sequences using CRISPRtarget²⁶. The
117 three sequenced strains of *G. st.* provided 77 spacer sequences, and 3 of them had
118 high-confidence viral and plasmid targets (Supplementary Fig. 2, Supplementary Table
119 3.). Extracting the sequences 3' of the targeted sequence revealed a consensus of 5'-
120 NNNNCNAA-3' (Fig. 2a, ED Fig. 2). Given the low number of viral targets, we next
121 performed cleavage assays on substrates containing various PAM sequences, revealing
122 a complete PAM sequence of 5'-NNNNCRAA-3' (Fig. 2b).

123 In addition to the CRISPR loci found in *G. st.* strains, we also found a type II
124 CRISPR locus in *Geobacillus* LC300 containing a Cas9 with ~97% amino acid identity to
125 the *G. st.* Cas9. Despite having nearly identical sequences, alignment of these two
126 homologs of GeoCas9 revealed a tight cluster of mutations in the PAM interacting
127 domain (PI). Furthermore, mapping these mutations onto the homology model of
128 GeoCas9 showed that they are located near the PAM region of the target DNA (Fig. 2c,
129 d). We hypothesized that this GeoCas9 variant might have evolved altered PAM
130 specificity. By searching for viral targets using the spacers in the *G. LC300* array, we
131 identified a preference for GMAA in place of the CRAA PAM of *G. st.*, lending support to
132 our hypothesis. We constructed and purified a hybrid Cas9 protein in which the PI
133 domain of the *G. LC300* Cas9 was substituted for the PI domain of *G. st.* Cas9 and
134 tested cleavage activity on targets containing various PAM sequences (Fig. 2b). We
135 found that, as predicted by protospacer sequences, the hybrid Cas9 preferred a GMAA
136 PAM rather than the CRAA PAM utilized by GeoCas9. Moreover, *G. LC300* appears to
137 be more specific for its optimal PAM, which may result in lower off-target cleavage for
138 genome editing applications²⁷. By creating a hybrid Cas9 with this naturally occurring
139 PAM-recognition variant, we double the sequence space that can be targeted by

140 GeoCas9 without resorting to structure based protein engineering as has been done for
141 other Cas9 homologs²⁷.

142

143 ***Identification of both crRNA and tracrRNA and engineering of GeoCas9 sgRNA***

144 CRISPR-Cas9 systems use a trans-activating crRNA (tracrRNA), which is
145 required for maturation of the crRNA and activation of Cas9^{6,28}. To identify the tracrRNA
146 for GeoCas9, we cultured *G. st.* and deep sequenced the small RNA it produced. We
147 found that the CRISPR array was transcribed despite a lack of phage or plasmid
148 challenge, and that the array was transcribed in the opposite direction of the Cas
149 proteins (Fig. 3a). The crRNA was processed to 23nt (Fig. 3b) of the spacer sequence
150 and 18nt of the repeat sequence *in vivo*, similar to other small type IIC Cas9 proteins^{8,29}.
151 Mapping of the RNAseq reads to the CRISPR array also revealed a putative tracrRNA
152 upstream of the Cas9 open reading frame (ORF).

153 We joined this putative tracrRNA to the processed crRNA using a GAAA-
154 tetraloop to generate a single-guide RNA (sgRNA)³⁰. Variations of this sgRNA were *in*
155 *vitro* transcribed and tested for their ability to direct GeoCas9 to cleave a radiolabeled
156 double-stranded DNA target at 37°C. We first varied the length of the crRNA:tracrRNA
157 duplex and found that this modification had little impact on the DNA cleavage rate (left
158 panel, Fig. 3c), making it a valuable place for further sgRNA modifications³¹. Next, we
159 tested the length of the tracrRNA, choosing stopping points near predicted rho-
160 independent terminators. In contrast to the crRNA:tracrRNA duplex length, the length of
161 the tracrRNA had a dramatic effect on the cleavage rate, with sequences shorter than
162 91nt supporting only a small amount of cleavage (middle panel, Fig. 3c). Finally, we
163 varied the length of the spacer sequence and found that 21–22nt resulted in a more than
164 5-fold increase in cleavage rate, compared to the 20nt spacer preferred by SpyCas9
165 (right panel, Fig. 3c). This finding contrasts with the most abundant spacer length of 23nt

166 found by RNAseq. This difference may be due to inter- or intramolecular guide
167 interactions in the *in vitro* transcribed sgRNA³².

168

169 ***Genome editing by GeoCas9 RNPs in mammalian cells***

170 With evidence that GeoCas9 maintains cleavage activity at mesophilic
171 temperatures, we assessed the ability of GeoCas9 to edit mammalian genomes. We
172 compared GeoCas9 and SpyCas9 editing efficiency by delivering preassembled
173 ribonucleoprotein complexes (RNPs) into cultured cells, circumventing differences
174 between SpyCas9 and GeoCas9 protein expression. First, GeoCas9 RNPs targeting
175 regions adjacent to various PAM sequences were delivered into HEK293T cells
176 expressing a destabilized GFP (Fig. 4a). We found that when targeted to sequences
177 adjacent to the preferred CRAA PAM, GeoCas9 decreased GFP fluorescence at levels
178 comparable to those observed for SpyCas9 (Fig. 4a). Next, we targeted GeoCas9 to
179 cleave the native genomic loci DNMT1 and AAVS1 (Fig. 4b,c). We varied the length of
180 the targeting spacer sequence and found that at one site 21nt was a sufficient length to
181 efficiently induce indels while at another site a 22nt spacer length was necessary. Given
182 this variability and that extending the spacer length to 22nt had no detrimental effects,
183 we conclude that a 22nt guide segment length is preferred for use in genome editing
184 applications. Moreover, when we tested editing efficiency at a site containing an
185 overlapping PAM for both GeoCas9 and SpyCas9, we observed similar editing
186 efficiencies by both proteins (Fig. 4b). At the DNMT1 locus we titrated amounts of
187 GeoCas9 and SpyCas9 RNPs to assess the effect on genome editing efficiency (Fig.
188 4c). Products analyzed by T7E1 assay again showed efficient production of indels by
189 both GeoCas9 and SpyCas9. These results demonstrate that GeoCas9 is an effective
190 alternative to SpyCas9 for genome editing in mammalian cells.

191

192 ***GeoCas9 functions over a wide temperature range and has extended lifetime in***
193 ***human plasma***

194 Based on initial observations showing that GeoCas9 protein remains folded at
195 elevated temperatures (Fig. 1d, e), we tested whether the GeoCas9 RNP maintains
196 activity after exposure to high temperatures. We incubated SpyCas9 and GeoCas9 at a
197 challenge temperature and added equimolar substrate to test the fraction of RNP that
198 remained functional. After incubation for 10 min at 45°C, the fraction of active SpyCas9
199 was greatly reduced (Fig. 5a). In contrast, the fraction of GeoCas9 after incubation at
200 45°C remained at 100% and not until challenge at 70°C did we detect a decrease in
201 activity (Fig. 5a).

202 Often thermostability comes at the cost of reduced activity at lower
203 temperatures³³. However, the wide range of natural growth temperatures for *G. st.*
204 suggested that GeoCas9 might maintain activity at mesophilic temperatures. To examine
205 this hypothesis, we measured the cleavage rate of SpyCas9 and GeoCas9 at various
206 temperatures (Fig. 5b). SpyCas9 DNA cleavage rates increased between 20–35°C,
207 reaching maximum levels from 35–45°C. Above these temperatures, SpyCas9 activity
208 dropped sharply to undetectable levels, as predicted by thermostability measurements.
209 In contrast, GeoCas9 activity increased to a maximum measured value at 50°C and
210 maintained maximum detectable activity up to 70°C, dropping to low levels at 75°C.
211 These results make GeoCas9 a valuable candidate for editing obligate thermophilic
212 organisms and for biochemical cleavage applications requiring Cas9 to operate at
213 elevated temperatures.

214 It was shown previously that thermostable proteins have longer lifetime in
215 blood³⁴. To test if this is the case for GeoCas9, we incubated SpyCas9 and GeoCas9 in
216 diluted human plasma at 37°C for 8 hrs and measured the amount of Cas9 activity
217 remaining (Fig. 5c). Although SpyCas9 maintained activity when incubated in reaction

218 buffer at 37°C, its activity was abolished even at the lowest concentration of plasma. In
219 contrast, GeoCas9 maintained robust activity after incubation with human plasma,
220 making it a promising candidate for *in vivo* RNP delivery.

221

222 **Discussion**

223 Our results establish GeoCas9 as a thermostable Cas9 homolog and expand the
224 temperatures at which Cas9 can be used. We anticipate that the development of
225 GeoCas9 will enhance the utility of CRISPR-Cas9 technology at both mesophilic and
226 thermophilic temperatures. The ability of Cas9 to function reliably in a wide range of
227 species has been key to its rapid adoption as a technology, but the previously developed
228 Cas9 homologs are limited for use in organisms that can grow below 42°C. The
229 complementary temperature range of GeoCas9 with SpyCas9 (Fig. 5c) opens up Cas9–
230 based genome editing to obligate thermophiles and facultative thermophiles, without the
231 additional steps of altering the temperature of the organisms. We also anticipate that
232 GeoCas9 will be useful for *in vitro* molecular biology applications requiring targeted
233 cleavage at elevated temperatures. Furthermore, we predict that the extended lifetime
234 of GeoCas9 in human plasma may enable more efficient delivery of Cas9 RNPs.

235 We were interested to note that GeoCas9 and SpyCas9 induced similar levels of
236 indels in HEK293T cells as SpyCas9 when delivered as an RNPs (Fig. 4b-d), despite
237 GeoCas9's lower DNA cleavage rate at 37°C (Fig. 5b). We conclude that biochemical
238 cleavage rates may not reflect limiting step of target search in a human cell. It may be
239 that GeoCas9 can persist longer in cells, which raises its effective concentration over
240 time and compensates for its slower cleavage rate. Moreover, in applications requiring
241 delivery of Cas9 into the bloodstream, the benefit of improved stability by GeoCas9 may
242 become even more apparent.

243 The development of GeoCas9 hinged upon utilizing the naturally occurring
244 diversity of CRISPR systems. The sheer abundance and diversity of Cas9 makes it
245 advantageous over newer type V systems, such as Cpf1, for developing specialized
246 genome editing tools. It has previously been suggested that type II CRISPR systems are
247 only found in mesophilic bacteria, and that protein engineering would be required to
248 develop a thermophilic Cas9³⁵. The rarity of type II CRISPR systems in thermophiles is
249 surprising given that CRISPR systems in general are enriched in thermophilic bacteria
250 and archaea³⁶. However, by searching the continually growing number of sequenced
251 bacteria, we uncovered a naturally occurring thermophilic Cas9. Exploiting Cas9
252 sequence diversity, rather than engineering thermostability, revealed a protein that
253 maintains activity over a broad temperature range, which is often difficult to select for
254 using directed evolution. Using the natural context of this CRISPR system, including the
255 transcribed RNA and targeted sequences, we further developed GeoCas9 with minimal
256 experimental optimization. The strategy of mining the natural context and diversity of
257 CRISPR systems has proven successful for uncovering novel interference proteins^{16,37},
258 and we anticipate that it can be applied more broadly to discover and develop new
259 genome editing tools.

260

261 References

- 262 1. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and
263 Beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
- 264 2. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and
265 evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105
266 (2014).
- 267 3. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and
268 evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105

- 269 (2014).
- 270 4. Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K. & Doudna, J. A. Single-
271 Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell* **60**,
272 398–407 (2015).
- 273 5. Ran, F. A. *et al.* In vivo genome editing using *Staphylococcus aureus* Cas9.
274 *Nature* **520**, 186–191 (2015).
- 275 6. Jinek, M. *et al.* A Programmable Dual-RNA – Guided. **337**, 816–822 (2012).
- 276 7. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and
277 editing. *Nat. Methods* **10**, 1116–21 (2013).
- 278 8. Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using
279 Cas9 from *Neisseria meningitidis*. *Proc. Natl. Acad. Sci.* **110**, 15644–15649
280 (2013).
- 281 9. Hirano, H. *et al.* Structure and Engineering of *Francisella novicida* Cas9. *Cell* **164**,
282 1–12 (2016).
- 283 10. Cong, L., Ran, F., Cox, D., Lin, S. & Barretto, R. Multiplex Genome Engineering
284 Using CRISPR / Cas Systems. *Science (80-.)*. (2013). doi:10.1038/nbt1319
- 285 11. Xiang, G., Zhang, X., An, C., Cheng, C. & Wang, H. Temperature effect on
286 CRISPR-Cas9 mediated genome editing. *Journal of Genetics and Genomics* 3–9
287 (2016). doi:10.1016/j.jgg.2017.03.004
- 288 12. Mougiakos, I. *et al.* Efficient genome editing of a facultative thermophile using the
289 mesophilic spCas9. *ACS Synth. Biol.* acssynbio.6b00339 (2017).
290 doi:10.1021/acssynbio.6b00339
- 291 13. Zeldes, B. M. *et al.* Extremely thermophilic microorganisms as metabolic
292 engineering platforms for production of fuels and industrial chemicals. *Front.*
293 *Microbiol.* **6**, 1–17 (2015).
- 294 14. Porteus, M. Genome Editing: A New Approach to Human Therapeutics. *Annu.*

- 295 *Rev. Pharmacol. Toxicol.* **56**, 163–190 (2016).
- 296 15. Staahl, B. T. *et al.* Efficient genome editing in the mouse brain by local delivery of
297 engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* (2017).
298 doi:10.1038/nbt.3806
- 299 16. Burstein, D., Harrington, L. B., Strutt, S. C. & Probst, A. J. New CRISPR-Cas
300 systems from uncultivated microbes. *Nat. Publ. Gr.* **542**, 237–241 (2017).
- 301 17. Markowitz, V. M. *et al.* IMG: The integrated microbial genomes database and
302 comparative analysis system. *Nucleic Acids Res.* **40**, 115–122 (2012).
- 303 18. Donk, P. J. A Highly Resistant Thermophilic Organism. *J. Bacteriol.* **5**, 373–374
304 (1920).
- 305 19. Notomi, T. *et al.* Loop-mediated isothermal amplification of DNA. *Nucleic Acids*
306 *Res.* **28**, E63 (2000).
- 307 20. Fujii, M., Takagi, M., Imanaka, T. & Aiba, S. Molecular Cloning of a Thermostable
308 Neutral Protease Gene from *Bacillus stearothermophilus* in a Vector Plasmid and
309 Its Expression in *Bacillus stearothermophilus* and *Bacillus subtilis*. *Microbiology*
310 **154**, 831–837 (1983).
- 311 21. Ingram, L. O. *et al.* Metabolic engineering for production of biorenewable fuels
312 and chemicals: Contributions of synthetic biology. *J. Biomed. Biotechnol.* **2010**,
313 (2010).
- 314 22. Cordova, L. T., Long, C. P., Venkataramanan, K. P. & Antoniewicz, M. R.
315 Complete genome sequence, metabolic model construction and phenotypic
316 characterization of *Geobacillus* LC300, an extremely thermophilic, fast growing,
317 xylose-utilizing bacterium. *Metab. Eng.* **32**, 74–81 (2015).
- 318 23. Mojica, F. J. M. *et al.* Short motif sequences determine the targets of the
319 prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
- 320 24. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly

- 321 interspaced short palindrome repeats (CRISPRs) have spacers of
322 extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
- 323 25. Paez-Espino, D. *et al.* CRISPR immunity drives rapid phage genome evolution in
324 streptococcus thermophilus. *MBio* **6**, 1–9 (2015).
- 325 26. Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. & Brown, C. M.
326 CRISPRTarget. *RNA Biol.* **10**, 817–827 (2013).
- 327 27. Kleinstiver, B. P. *et al.* Engineered {CRISPR-Cas9} nucleases with altered {PAM}
328 specificities. *Nature* **523**, 481–485 (2015).
- 329 28. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and
330 host factor RNase III. *Nature* **471**, 602–607 (2011).
- 331 29. Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-
332 RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids*
333 *Res.* **42**, 2577–2590 (2014).
- 334 30. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in
335 Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–822 (2012).
- 336 31. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered
337 CRISPR-Cas9 complex. *Nature* **517**, 583–8 (2014).
- 338 32. Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E. & Schier, A. F. Internal
339 guide RNA interactions interfere with Cas9-mediated cleavage. *Nat. Commun.* **7**,
340 11750 (2016).
- 341 33. Sawle, L. & Ghosh, K. How do thermophilic proteins and proteomes withstand
342 high temperature? *Biophys. J.* **101**, 217–227 (2011).
- 343 34. Narasimhan, D. *et al.* Structural analysis of thermostabilizing mutations of cocaine
344 esterase. *Protein Eng. Des. Sel.* **23**, 537–547 (2010).
- 345 35. Li, Y. *et al.* Harnessing Type I and Type III CRISPR-Cas systems for genome
346 editing. *Nucleic Acids Res.* **44**, (2015).

- 347 36. Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S. & Koonin, E. V.
348 Viral diversity threshold for adaptive immunity in prokaryotes. *MBio* **3**, 1–10
349 (2012).
- 350 37. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nat.*
351 *Rev. Microbiol.* **15**, 169–182 (2017).
- 352 38. Wright, A. V *et al.* Rational design of a split-Cas9 enzyme complex. *Proc. Natl.*
353 *Acad. Sci. U. S. A.* **112**, 2984–9 (2015).
- 354 39. Zhang, Y. *et al.* Processing-Independent CRISPR RNAs Limit Natural
355 Transformation in *Neisseria meningitidis*. *Mol. Cell* **50**, 488–503 (2013).
- 356 40. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
357 *Methods* **9**, 357–359 (2012).
- 358 41. Lin, S., Staahl, B., Alla, R. K. & Doudna, J. a. Enhanced homology-directed
359 human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*
360 **3**, 1–13 (2014).
- 361 42. Kelly, L. A., Mezulis, S., Yates, C., Wass, M. & Sternberg, M. The Phyre2 web
362 portal for protein modelling, prediction, and analysis. *Nat. Protoc.* **10**, 845–858
363 (2015).

364

365 **Methods:**

366 **Identification of thermophilic Cas9 homologs and generation of heterologous**
367 **expression plasmids.** We mined all isolate genomes from the public Integrated
368 Microbial genomes (IMG) database¹⁷ using the “Genome Search by Metadata Category
369 tool.” We selected all the genomes annotated as “thermophile” (336) or
370 “hyperthermophile” (94) and searched for the presence of Cas9-like candidates (hits to a
371 TIGRfam model 01865 for Csn-like or 03031 for a Csx12-like) contained within a full
372 CRISPR-Cas system (presence of Cas1, Cas2, and a Repeat-Spacers array). We

373 initially selected the GeoCas9 variant due to its completeness, smaller gene size
374 (shorter than the widely used SpyCas9), and growth in a large temperature range from
375 30-75 (optimal at 55°C). The Cas9 from *Geobacillus stearothermophilus* was codon
376 optimized for *E. coli*, ordered as Gblocks (IDT) and assembled using Gibson Assembly.
377 For protein expression, a pET based plasmid containing an N terminal 10xHis-tag and
378 MBP was used. For PAM depletion assays, a p15A plasmid was generated with the
379 sgRNA constitutively expressed. The plasmids used are available from Addgene and
380 their maps can be found in Supp. Table 2.

381

382 **Cas9 purification.** Cas9 was purified as previously described⁶ with modification. After
383 induction, *E. coli* BL21(DE3) expressing Cas9 was grown in Terrific Broth overnight at
384 18°C. Cells were harvested, re-suspended in Lysis Buffer (50mM Tris-HCl, pH 7.5,
385 20mM imidazole, 0.5mM TCEP, 500mM NaCl, 1mM PMSF), broken by sonication, and
386 purified on Ni-NTA resin. TEV was added to the elution and allowed to cleave overnight
387 at 4°C. The resulting protein was loaded over tandem columns of an MBP affinity column
388 onto a heparin column and eluted with a linear gradient from 300mM to 1250mM NaCl.
389 The resulting fractions containing Cas9 were purified by gel filtration chromatography
390 and flash frozen in Storage Buffer (20mM HEPES-NaOH pH 7.5, 5% Glycerol, 150mM
391 NaCl, 1mM TCEP).

392

393 **Differential Scanning Calorimetry.** Cas9 proteins were dialyzed into degassed DSC
394 Buffer (0.5mM TCEP, 50mM KH₂PO₄(pH 7.5), 150mM NaCl) overnight at 4°C. Samples
395 were diluted to 0.3mg/ml and loaded a sample cell of a NanoDSC (TA instruments);
396 buffer alone was used in the reference cell. The cell was pressurized to 3atm and the
397 sample was heated from 20 to 90°C. Measurements made for buffer in both the sample
398 and reference cells were subtracted from the sample measurements.

399

400 **Biochemical cleavage assays.** Radioactive cleavage assays were conducted as
401 previously described³⁸. Reactions were carried out in 1× Reaction Buffer (20mM Tris-
402 HCl, pH 7.5, 100mM KCl, 5mM MgCl₂, 1mM DTT and 5% glycerol (v/v)). 100nM Cas9
403 and 125nM sgRNA were allowed to complex for 5min at 37°C. ~1nM radiolabeled probe
404 was added to the RNP to initiate the reaction. Finally, the reaction was quenched with 2×
405 Loading Buffer (90% formamide, 20mM EDTA, 0.02% bromophenol blue, 0.02% xylene
406 cyanol and products were analyzed on 10% urea-PAGE gel containing 7M urea.

407 For thermostability measurements (Fig. 4a), 100nM Cas9 was complexed with
408 150nM sgRNA in 1x Reaction Buffer for 5min at 37°C. 100nM of a PCR product
409 containing the targeted sequence was cleaved using dilutions of the estimated 100nM
410 RNP complex to accurately determine a 1:1 ratio of Cas9 to target. Next, samples were
411 challenged at the indicated temperature (40°C–75°C) for 10min and then returned to
412 37°C. 100nM PCR product containing the targeted sequence was added to the reaction
413 and it was allowed to react for 30min at 37°C. The reaction was quenched with 6×
414 Quench Buffer (15% glycerol (v/v), 1mg/ml Orange G, 100mM EDTA) and products were
415 analyzed on a 1.25% agarose gel stained with ethidium bromide.

416 For thermophilicity measurements (Fig. 4b), 500nM Cas9 was complexed with
417 750nM sgRNA in 1x Reaction Buffer for 5min at 37°C. The samples were placed at the
418 assayed temperature (20°C–80°C) and 100nM of PCR product was added to the
419 reaction. Time points were quenched using 6× Quench Buffer and analyzed on a 1.25%
420 agarose gel stained with SYBR Safe (Thermo Fisher Scientific).

421 To study the effect of human plasma on the stability of Cas9 proteins,
422 preassembled Cas9-RNP was incubated for 8 hours either at 37°C or 4°C in Reaction
423 Buffer with the specified amount of normal human plasma. Substrate was then added
424 and cleavage products were analyzed as described for thermostability measurements.

425

426 **Small RNA sequencing.** *Geobacillus stearothermophilus* was obtained from ATCC and
427 cultured at 55°C in Nutrient Broth (3g beef extract and 5g peptone per liter water, pH
428 6.8) to saturation. Cells were pelleted and RNA was extracted using a hot phenol
429 extraction as previously described³⁹. Total RNA was treated with TURBO DNase
430 (Thermo Fisher Scientific), rSAP (NEB) and T4 PNK (NEB) according to manufactures
431 instructions. Adapters were ligated onto the 3' and 5' ends of the small RNA, followed by
432 reverse transcription with Superscript III. The library was amplified with limited cycles of
433 PCR, gel-extracted on an 8% native PAGE gel and sequenced on an Illumina MiSeq.
434 Adapters were trimmed using Cutadapt and sequences >10nt were mapped to the *G. st.*
435 CRISPR locus using Bowtie 2⁴⁰.

436

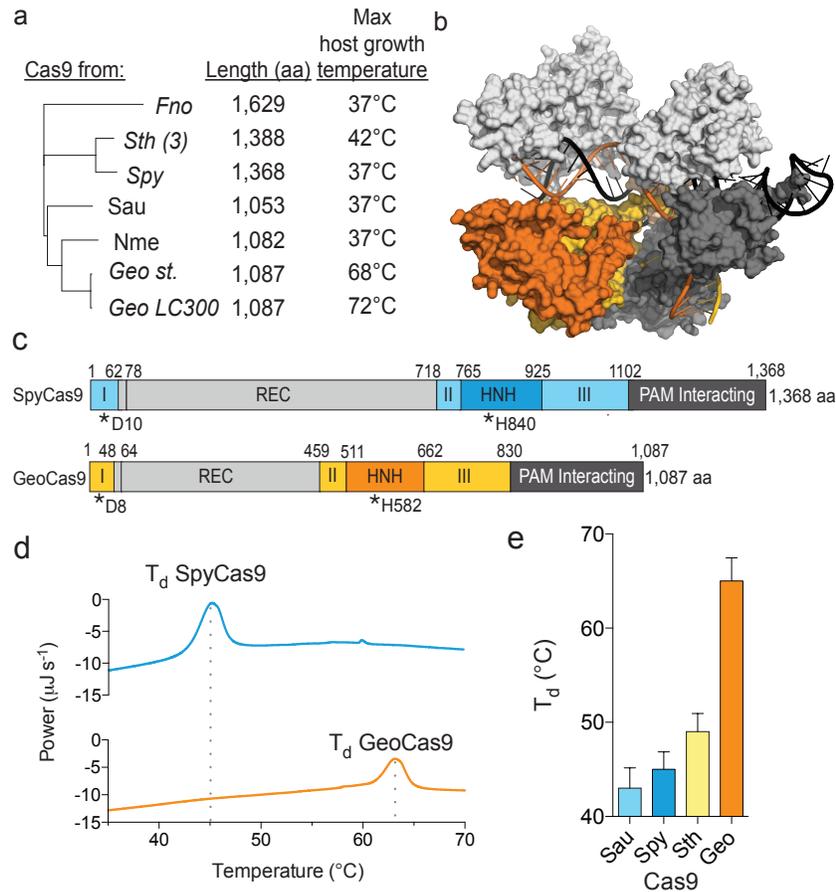
437 **HEK293T EGFP disruption assay and indel analysis.** HEK293T cells expressing a
438 destabilized GFP were grown in Dulbecco's Modification of Eagle's Medium (DMEM)
439 with 4.5g/L glucose L-glutamine and sodium pyruvate (Corning Cellgro), supplemented
440 with 10% fetal bovine serum, penicillin and streptomycin at 37°C with 5% CO₂. ~24hrs
441 before transfection, $\sim 3 \times 10^4$ cells were seeded into each well of a 96-well plate. The
442 next day, 20pmol (unless otherwise specified) of RNP was assembled as previously
443 described⁴¹ and mixed with 10 μ L OMEM. The RNP was added to 10 μ L of 1:10 dilution of
444 Lipofectamine 2000 (Life Technologies) in OMEM and allowed to incubate at room
445 temperature for 10min and added to the cells. Cells were analyzed for GFP fluorescence
446 48h later using Guava EasyCyte 6HT. Experiments were conducted in triplicate and the
447 mean \pm S.D. is shown. For analysis of indels, genomic DNA was extracted using Quick
448 Extraction Solution (Epicentre), and the DNMT1 and AAVS1 loci were amplified by PCR.
449 T7E1 reaction was conducted according to the manufacturer's instructions and products

450 were analyzed on a 1.5% agarose gel stained with SYBR gold (Thermo Fisher

451 Scientific).

452

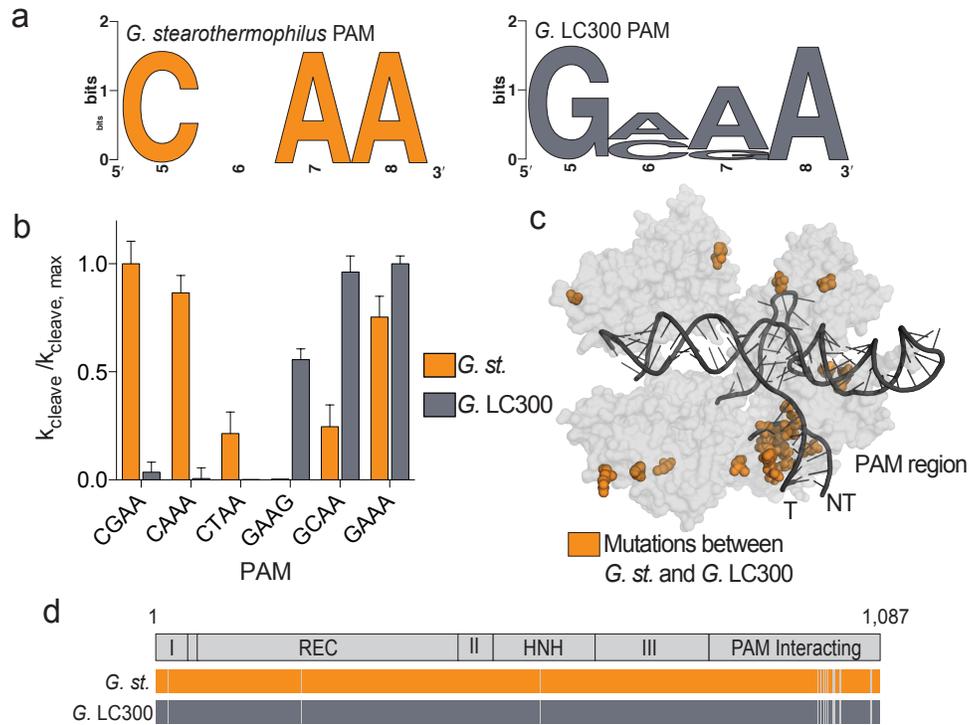
453 **Figures**



454

455 **Figure 1 | GeoCas9 is a small, thermostable Cas9 homolog.** **a**, Phylogeny of Cas9
 456 proteins used for genome editing with their length (amino acids) and the maximum
 457 temperatures that supports growth of the host indicated to the right²² (*Nme*, *Neisseria*
 458 *meningitidis*; *Geo*, *Geobacillus stearothermophilus*; *Geo LC300*, *Geobacillus LC300*;
 459 *Spy*, *Streptococcus pyogenes*; *Sau*, *Streptococcus aureus*; *Fno*, *Francisella novicida*;
 460 *Sth* (3), *Streptococcus thermophilus* CRISPR III). **b**, Homology model of GeoCas9
 461 generated using Phyre²⁴² with the DNA from PDB 5CZZ docked in. **c**, Schematic
 462 illustration of the domains of Spy Cas9 (blue) and GeoCas9 (orange) with active site
 463 residues indicated below with asterisks. **d**, Representative traces for Differential
 464 Scanning Calorimetry (DSC) of GeoCas9 and SpyCas9, T_d ; Denaturation temperature.

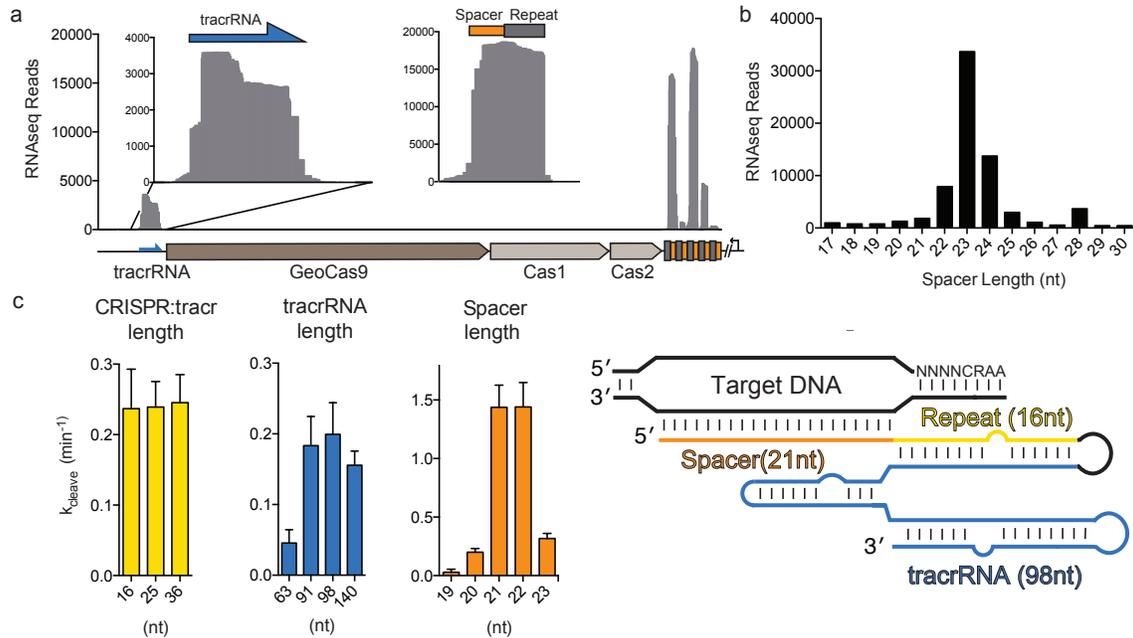
465 **e**, Denaturation temperature of various Cas9 proteins as measured by DSC, mean \pm
466 S.D. is shown.



467

468 **Figure 2 | PAM identification and engineering of GeoCas9. a**, WebLogo for
469 sequences found at the 3' end of protospacer targets identified with CRISPRTarget for
470 *Geobacillus stearotherophilus* (left panel) and *Geobacillus* LC300 (right panel). **b**,
471 Cleavage assays conducted with the two homologs of GeoCas9. Substrates with
472 various PAM sequences were P32-labelled and mean \pm S.D. is shown. **c**, Mapping of
473 mutated residues (orange spheres) between *G. st.* and *G. LC300* onto the homology
474 model of GeoCas9 showing high density in the PAM interacting domain near the PAM
475 region of the target DNA. **d**, Alignment of the Cas9 proteins from *G. st.* and *G. LC300*
476 with the domain boundaries shown above. Solid colors represent identical residues and
477 grey lines indicate residues that are mutated between the two Cas9 homologs.

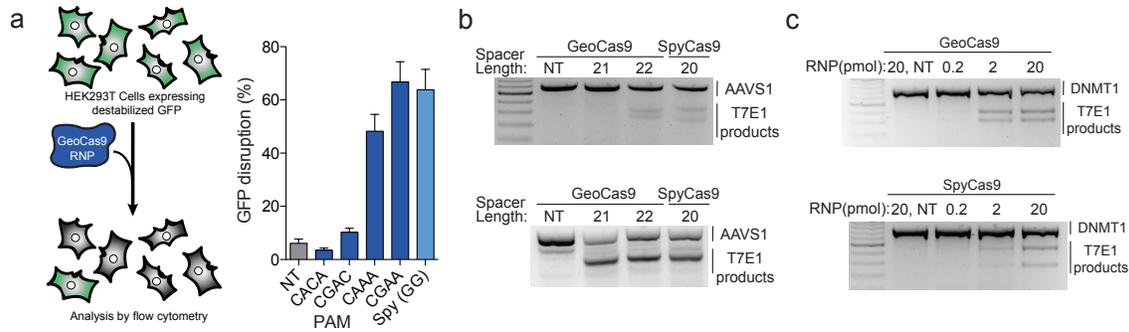
478



479

480 **Figure 3 | Small RNA-seq and sgRNA engineering for GeoCas9.** **a**, Small RNA
 481 sequenced from *G. stearothermophilus* mapped to the CRISPR locus. Inset shows
 482 enlargement of the region corresponding to the tracrRNA and the most highly
 483 transcribed repeat and spacer sequence. **b**, Distribution of the length of the spacer
 484 sequences extracted from the small RNA sequencing results. **c**, Length optimization of
 485 the tracrRNA and crRNA for GeoCas9 and the optimal guide RNA design (right panel).
 486 The length of the tracrRNA, crRNA:tracrRNA duplex and spacer was optimized
 487 sequentially by transcribing variations of the sgRNA and testing their ability to guide
 488 GeoCas9-mediated cleavage of a radiolabeled substrate. The mean $k_{\text{cleave}} \pm$ S.D. is
 489 shown and experiments were conducted in triplicate.

490



491

492 **Figure 4 | Genome editing activity of GeoCas9 in mammalian cells. a, EGFP**

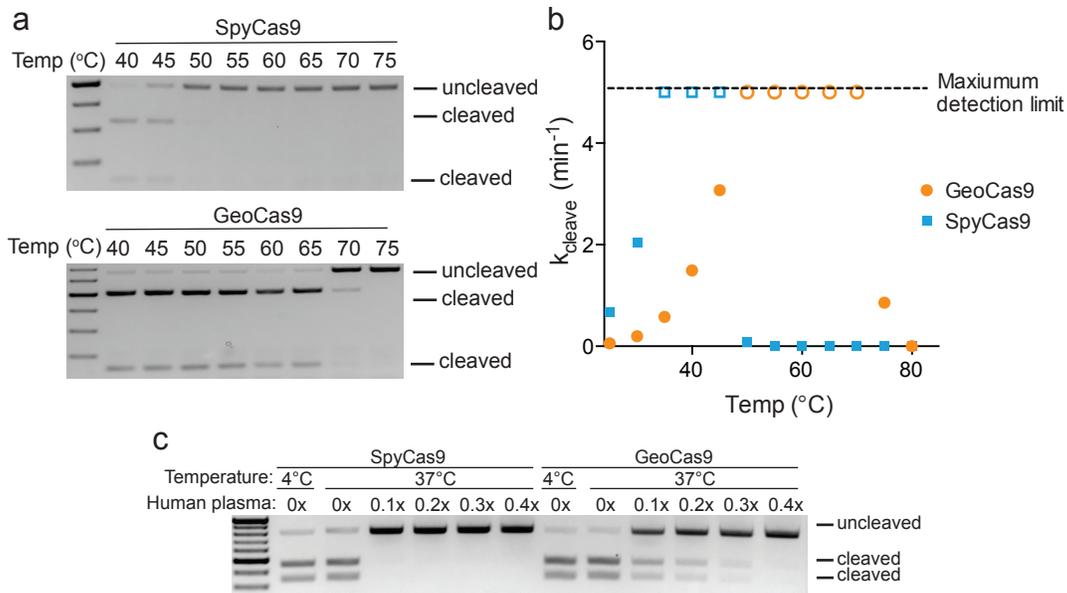
493 disruption in HEK293T cells by GeoCas9. HEK293FT cells expressing a destabilized
494 GFP were transfected with GeoCas9 RNP preassembled with a targeting or non-
495 targeting guide RNA. Cells were analyzed by flow cytometry and targets adjacent
496 to the CRAA PAM resulted in efficient GFP disruption (NT; non-targeting). **b,**

497 T7E1 analysis of indels produced at the AAVS1 locus when the guide length was

498 varied from 21 to 22nt. **c,** T7E1 analysis of indels produced using a titration of

499 GeoCas9 and SpyCas9 RNP targeting the DNMT1 locus in HEK293T cells.

500



501

502 **Figure 5 | Thermostability of GeoCas9 and longevity in human plasma. a**, Activity of

503 SpyCas9 and GeoCas9 after incubation at the indicated temperature. After challenging

504 at the higher temperature, reactions were conducted at 37°C using a 1:1 ratio of

505 substrate to RNP. b, Cleavage rate of SpyCas9 and GeoCas9 RNPs at various

506 temperatures. Maximum detection limit is shown by the dashed line at $k_{\text{cleave}}=5$,

507 indicating that the reaction completed in ≤ 30 s. c, Effect of incubating GeoCas9 and

508 SpyCas9 in human plasma. After incubation in varying concentrations of human plasma

509 for 8hrs at 37°C, the reaction was carried out with 1:1 ratio of DNA substrate to RNP.

510 **Supplementary Table 1 | DNA and RNA sequences used in this study**

511 **Supplementary Table 2 | Thermophilic Cas9 candidates**

512 **Supplementary Table 3 | Spacers identified from *Geobacillus stearothermophilus***

513 **and *Geobacillus* LC300**

514

515 **Acknowledgments**

516 We thank A. Tambe, N. Ma and K. Zhou for technical assistance and discussions. L.B.H

517 and J.S.C are supported by a US National Science Foundation Graduate Research

518 Fellowship. J.A.D. is an Investigator of the Howard Hughes Medical Institute. This

519 research was supported in part by the Allen Distinguished Investigator Program, through

520 The Paul G. Allen Frontiers Group and the National Science Foundation (MCB-1244557

521 to J.A.D.).

522

523 **Author Contributions**

524 L.B.H. designed and executed experiments with help from J.S.C., E.M., B.S and J.A.D.;

525 The search for thermophilic Cas9 homologs was conducted by D.P.E. and N. C. K. All

526 authors revised and agreed to the manuscript.

527

528 **Materials & Correspondence.** Plasmids used in this study are available on Addgene

529 (#87700, #87703). RNA sequencing data will be available on the NCBI Sequence Read

530 Archive (SRA). Correspondence and requests for materials should be addressed to

531 J.A.D. (doudna@berkeley.edu).

532

533 **Competing financial interests**

534 J.A.D. is executive director of the Innovative Genomics Institute at the University of

535 California, Berkeley (UC Berkeley) and the University of California, San Francisco

536 (UCSF). J.A.D. is a co-founder of Editas Medicine, Intellia Therapeutics and Caribou
537 Biosciences and a scientific adviser to Caribou, Intellia, eFFECTOR Therapeutics and
538 Driver. Funding has been received from the Howard Hughes Medical Institute (HHMI),
539 the US National Institutes of Health, the US National Science Foundation, Roche, Pfizer,
540 the Paul Allen Institute and the Keck Foundation. J.A.D. is employed by HHMI and works
541 at the UC Berkeley. UC Berkeley and HHMI have patents pending for CRISPR
542 technologies including this work on which the authors are inventors.
543